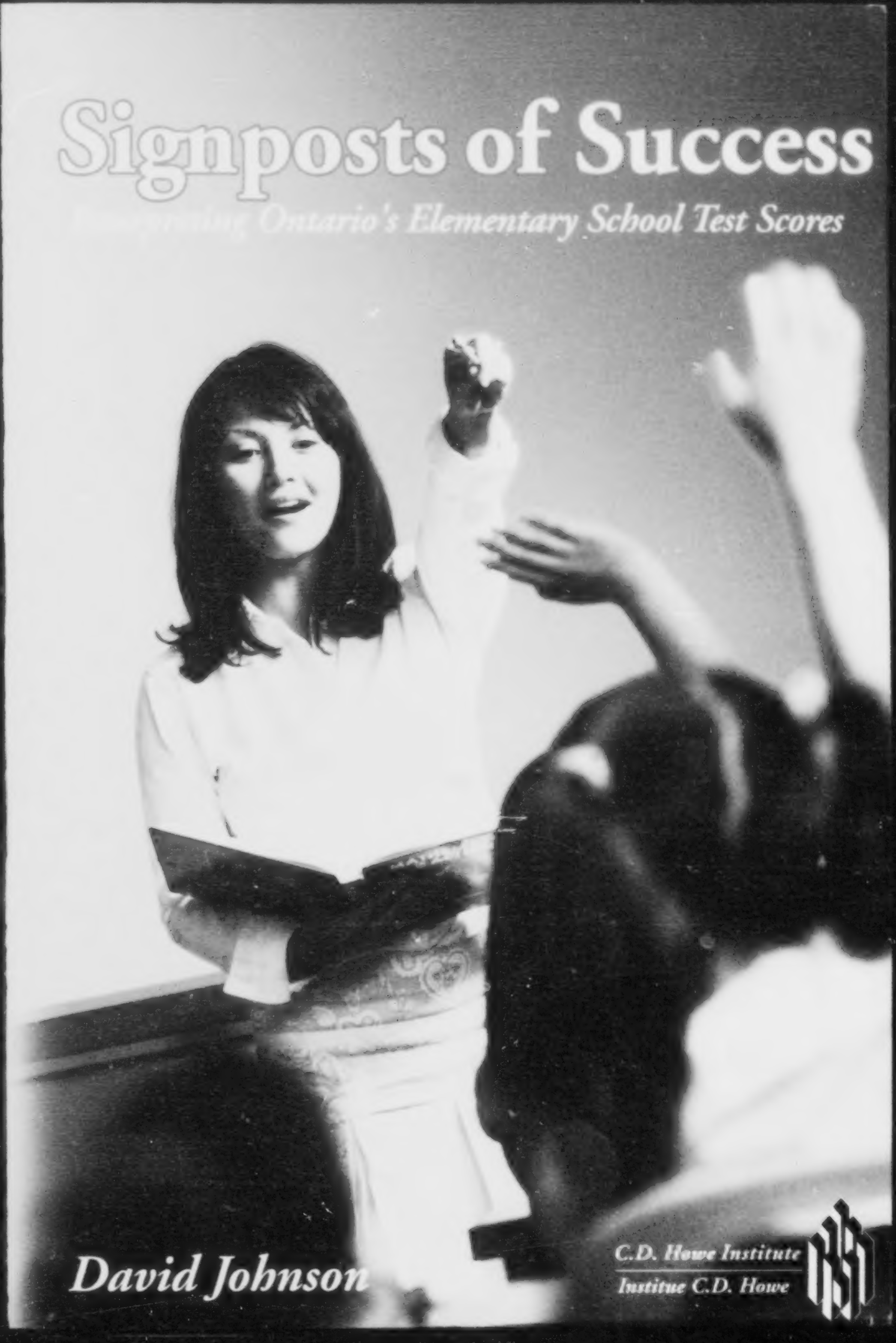
Signposts of Success

Interpreting Ontario's Elementary School Test Scores



David Johnson

C.D. Howe Institute

Institue C.D. Howe

The C.D. Howe Institute

The C.D. Howe Institute is a national, nonpartisan, nonprofit research organization that aims to improve Canadians' standard of living by fostering sound economic and social policy.

The Institute promotes the application of independent research and analysis to major economic and social issues affecting the quality of life of Canadians in all regions of the country. It takes a global perspective by considering the impact of international factors on Canada and bringing insights from other jurisdictions to the discussion of Canadian public policy. Policy recommendations in the Institute's publications are founded on quality research conducted by leading experts and subject to rigorous peer review. The Institute communicates clearly the analysis and recommendations arising from its work to the general public, the media, academia, experts, and policymakers.

The Institute began life in 1958 when a group of prominent business and labour leaders organized the Private Planning Association of Canada to research and promote educational activities on issues related to public economic and social policy. The PPAC renamed itself the C.D. Howe Research Institute in 1973 following a merger with the C.D. Howe Memorial Foundation, an organization created in 1961 to memorialize the Right Honourable Clarence Decatur Howe. In 1981, the Institute adopted its current name after the Memorial Foundation again became a separate entity in order to focus its work more directly on memorializing C.D. Howe. The C.D. Howe Institute will celebrate its 50th Anniversary as the gold standard for public policy research in 2008.

The Institute encourages participation in and support of its activities from business, organized labour, associations, the professions, and interested individuals. For further information, please contact the Institute's Development Officer.

The Chairman of the Institute is Guy Savard; Jack M. Mintz is President and Chief Executive Officer.

C.D. Howe Institute
67 Yonge Street
Suite 300
Toronto, Ontario M5E 1J8
tel.: 416-865-1904; fax: 416-865-1866
e-mail: cdhowe@cdhowe.org
Internet: www.cdhowe.org

105-03464/

Signposts of Success

Interpreting Ontario's Elementary School Test Scores

David Johnson

Policy Study 40

C.D. Howe Institute

C.D. Howe Institute publications are available from:
Renouf Publishing Company Limited, 5369 Canotek Road, Unit 1,
Ottawa, Ontario K1J 9J3
phone: (613) 745-2665; fax: (613) 745-7660;
Internet: www.renoufbooks.com

and from Renouf's store at:
71½ Sparks Street, Ottawa (613) 238-8985

This book is printed on recycled, acid-free paper.

Library and Archives Canada Cataloguing in Publication

Johnson, David R., 1956–

Signposts of success : interpreting Ontario's elementary school test scores / David Johnson.

(Policy study ; 40)
Includes bibliographical references.
ISBN 0-88806-649-X

1. Elementary schools — Ontario — Evaluation. 2. Education, Elementary — Ontario — Evaluation. 3. Educational tests and measurements — Ontario. I. C.D. Howe Institute. II. Title. III. Series: Policy study (C.D. Howe Institute) ; 40.

LB3054.C3J63 2005 372.9713 C2005-901379-6

© C.D. Howe Institute, Toronto.
Quotation with appropriate credit is permissible.
Cover design by Allison Klus.
Printed in Canada by Ricoh, 205 Industrial Parkway North, Aurora ON L4G 4C4
March 2005.

I have had wonderful teachers in my life — people with whom I enjoyed learning. Their influence on me continues. Some are no longer alive except in the memories of their students. I bring to my mind one special teacher (there was often more than one) from each school I have attended, and mention by name Miss Hazel Curry (Binkley Public School); Mr. Don Thomas (Dalewood Public School); Mrs. Adele Trussler (Westdale Secondary School); Mr. Bill Lawrence (London Central Secondary School); Professor John Munro (University of Toronto); Professor David Laidler (University of Western Ontario); and Professor Olivier Blanchard (Harvard University and MIT). People who teach well have a lasting influence on their students. A book about teaching should be dedicated to good teachers. This one is.

Contents

Foreword ix

Acknowledgments xi

1 *Can Provincewide Testing Identify Successful Schools?* 1
- Toward a Fair and Useful Comparison 3
- A Summary of the Findings 5
- A Guide to the Rest of the Study 10

2 *Ontario's School Assessments: Background Issues* 13
- EQAO's Activities 14
- Reporting the Assessment Results 16
- Using the Assessment Results 17
- Similar Studies in Other Jurisdictions 20
- Political and Institutional Reactions to the Ontario School Assessment Process 23
- How Much Does the Assessment Process Cost? 26
- Conclusion 28
- *Chapter Summary* 29

3 *Describing an Elementary School's Community* 31
- Measuring School Community Characteristics: A General Description 32
- What the Census and Postal Code Data Say about School Community Characteristics 33
- The EQUIP Data 51
- *Chapter Summary* 54
- Technical Appendix: Constructing the School Communities' Socio-Economic Indicators 56

4 *Analyzing Test Results for Ontario's Elementary Schools* 61
- The Structure of the Assessment Data 62
- A First Look at the Achievement Results 64
- Identifying School Community Variables Associated with High Test Results 70
- "Cheap and Easy" Sets of Context Variables 82
- Conclusions 84
- *Chapter Summary* 85
- Technical Appendix 88

5 *An Analysis of the Exemption Process* 116
A Broad Picture of Exemptions 119
Do School Boards Have Different Exemption Rates? 121
Chapter Summary 127
Technical Appendix 129

6 *Ranking Schools Using the EQAO Assessment Data* 131
A Reminder of the Variables Used in the Analysis 133
Creating a Measure of the School's Relative Performance 135
Ranking Schools: The Waterloo Region District School Board 148
Conclusions 165
Chapter Summary 165
Technical Appendix 169

7 *Listening to Principals, Teachers, and Parents at "Good" Schools* 180
Finding and Visiting "Good" Schools 180
What Principals Told Me 186
What Teachers Told Me 191
What Parents Told me 203
Summing Up the School Visits 206
Chapter Summary 206

8 *Conclusions* 209
Lessons from the Statistical Analysis 210
What I Heard at Visits to "Good" Schools 211
My Views of the Effectiveness of the Assessment Process 212
Recommendations for Improving the EQAO Process 217
A Final Note 224

References 225

Figures

3.1 Immigrants in Ontario Elementary School Communities 37
3.2 Language in Ontario Elementary School Communities 38
3.3 Aboriginals in Ontario Elementary School Communities 39
3.4 Education Levels of Adults in Ontario Elementary School Communities 40
3.5 The Relationship between the Education Variables 41
3.6 Single-Parent Families in Ontario Elementary School Communities 42
3.7 Unemployment in Ontario Elementary School Communities 43
3.8 Average Household Income in Households with Two or More People, Ontario Elementary School Communities 44
3.9 The Relationship between Average Household Income and Single-Parent Families 45

3.10 Measures of Mobility in Ontario Elementary School Communities 46
3.11 Detached Homes in Ontario Elementary School Communities 47
3.12 Enrollment in Ontario Elementary Schools 48
4.1 Trends in Average Achievement on Assessments, academic years 1997–98 to 2000-01 67
A4.1 A Positive Relationship between Assessment Results and Factor X 89
A4.2 A Negative Relationship between Assessment Results and Factor X 90
A4.3 The Relationship between Factor X and Two Assessment Results............91
A4.4 Two Assessments Where the Relationship to Factor X Differs 92
5.1 Average Exemption Rates by Assessment and Year, academic years 1998–99 to 2001-02 121
5.2 Distribution of Student Exemptions from Assessments..........................126
6.1 Difference between Actual and Predicted Adjusted Assessment Scores in Ontario Elementary Schools 144
6.2 A Visual Explanation of Relative School Rankings 146
6.3 The Relationship between Fraser Institute Ratings and Students Achieving Assessment Levels 3 and 4 in the Waterloo Region District School Board 158
6.4 The Relationship between Fraser Institute Ratings and Average Household Income in the Waterloo Region District School Board 160
6.5 The Relationship between Z-scores and Average Household Income in the Waterloo Region District School Board 162
6.6 The Relationship between Fraser Institute Ratings and Parental Educational Achievement in the Waterloo Region District School Board 163
6.7 The Relationship between Z-scores and Parental Education in the Waterloo Region District School Board 164
A6.1 The Relationship between Predicted Adjusted Scores and Actual Adjusted Scores, Grades 3 and 6 172

Tables

3.1 Coverage of Postal Code and Census Data for the Ontario Elementary School System 34
3.2 Participation in Postal Code Collection, academic year 2000-01 35
3.3 Descriptive Statistics for Schools in the Postal Code Sample 36
3.4 Measures of Income 44
3.5 Measures of Income, Correlation Index 44
3.6 Correlation Coefficients among 16 Variables that Describe Ontario Elementary School Communities 50
3.7 The Elementary School EQUIP Variables 53
4.1 Students in Each Assessment Data Category 66
4.2 The Association of School Community Socio-Economic Variables and Grade 3 and Grade 6 Assessment Results 76
A4.1 Variation in Assessment Results across Years 94
A4.2 Variation in Assessment Results across Tests..99
A4.3 The Association of Socio-Economic School Community Variables with Grade 3 Assessment Scores....................................102–03
A4.4 The Association of Socio-Economic School Community Variables with Grade 6 Assessment Scores....................................104–05

A4.5 An Interaction Model of Achievement in Grade 3 and Grade 6 106
A4.6 EQAO Context Variables as Predictors of Results 113
A4.7 The Association of Assessment Results with a Small Subset of Census Variables 114
5.1 The Exemption Process, academic years 1998–99 to 2001–02 120
5.2 Socio-Economic Determinants of Exemption Rates at Schools and Testing for Board Effects on Exemption Rates 123
6.1 Socio-Economic Characteristics of Schools in the Waterloo Region District School Board 150–53
6.2 Measures Used to Rank Schools in the Waterloo Region District School Board 154–57
A6.1 The Explanation of Adjusted School Scores 170
A6.2 Regressions Estimated in Figures 6.3, 6.4, 6.5, 6.6, and 6.7 174
A6.3 Examples of Fraser Institute "Value-Added" Regressions 175
A6.4 A Comparison of the Z-scores and Fraser Institute "Value-Added" Grades for 60 Schools in the Waterloo Region District School Board 176–77
7.1 Socio-Economic Characteristics of 13 "Good" Schools Visited 182–83
7.2 Why Teachers Thought Their School Performed Well 194
7.3 Activities Teachers Identified as Increasing EQAO Assessment Scores 195
7.4 Activities Teachers Identified as Not Emphasized in Favour of Improving EQAO Assessment Scores 196

About the Author 225

Members of the C.D. Howe Institute 227

Foreword

In recent years, provincial governments have developed new approaches to improve accountability in their school systems. One method has been to develop provincewide standardized tests to assess the progress of students in key subject areas in the hope of shedding some light on how schools are faring in their primary mission of educating their students. Ontario introduced standardized testing for elementary schools in the 1996–97 academic year for reading, writing, and mathematics.

Standardized testing is, however, controversial. Critics argue that school rankings are based, not on schools' relative success in educating their students, but on school community socio-economic factors — such as parents' educational background and family income — that influence students' performance. Thus, some schools may be ranked unfairly low simply because they draw their students from poor neighbourhoods.

In this book, David Johnson makes an important contribution to the literature on standardized testing by separating socio-economic factors from other factors that explain students' performance. Professor Johnson develops a methodology to rank schools according to the degree to which their students' test scores vary from what would be predicted for schools in communities with similar socio-economic characteristics. This methodology permits the identification of schools whose students perform better than expected, allowing educators to look more carefully at other factors that might explain good performance — such as the principal's managerial ability, the quality of teaching, and the availablity of resources. Educators will no doubt greet this approach as good news: they *can* make a difference in the classroom.

The C.D. Howe Institute hopes this book will open up a broad debate about standardized testing. Professor Johnson's recommendations for improving the current approach to testing and the collection of useful data should also help schools enhance their students' learning experience.

The Institute is grateful to the Donner Canadian Foundation for its assistance with funding for the publication of this study. We also wish to thank Bill Robson and Yvan Guillemette for coordinating and reviewing the research, Barry A. Norris for editing the manuscript and preparing it for publication, and Wendy Longsworth and Kevin Doyle for their production assistance.

As with all C.D. Howe Institute publications, the analysis and opinions presented here are the responsibility of the author and do not necessarily reflect the views of the Institute's members or Board of Directors.

Jack M. Mintz
President and
Chief Executive Officer

Acknowledgments

The list of people who have contributed to this project is long — appropriately, given its magnitude.

Tommy Wu has been a very helpful research assistant. Greg Lang and Megan Barr also provided excellent research assistance on the project. Jennifer Dickson, a former student in the Master in Arts, Business Economics at Wilfrid Laurier University, and I worked together on a pilot study of school assessment data with a small sample of schools in the Waterloo Regional District School Board. That work began this project.

Statistical assistance and advice were provided by Mike Veall, Terry Levesque, Tom Crossley, Rick Elliot, and the Statistical Consulting Service at the University of Waterloo. Howard Kim at the Ontario Ministry of Education and Kirsten McKnight at the Education Quality Accountability Office helped enormously in understanding and sorting the school files. Bo Wandscheider and Michelle Edwards at the University of Guelph's Tri-University Data Centre were invaluable in helping to handle the census files. Teachers and principals as well as staff at boards were helpful in making the school visits a success.

I would also like to acknowledge the assistance and helpful suggestions of the reviewers of this study, as well as those of my editor Barry A. Norris and the staff of the C.D. Howe Institute — particularly Jack Mintz, Bill Robson, and Yvan Guillemette.

Unavoidably, while writing this book, I thought about my own schooling, both its strengths and its weaknesses. Researching school testing touches many emotional nerves in Ontario. I am grateful for the patience shown by a variety of people. We all gingerly stepped through the minefield and worked together to try to understand the educational process better. There are no simple answers. I do not pretend to have the complete story — just a part of the picture. The errors in the analysis remain my responsibility, as are the conclusions and policy recommendations. The evidence is, however, inescapable: teachers and principals in local schools

make a difference in the average test scores of their students — a difference, moreover, that is not associated with the socio-economic characteristics of the school community.

David Johnson
January 2005

1

Can Provincewide Testing Identify Successful Schools?

Tests measure what students have learned. In elementary schools, students are assessed by their classroom teacher on their performance in reading, writing, and mathematics. These assessments identify the students' areas of strength and weakness. Everyone — parents, students, and teachers alike — takes such assessments for granted. Students take home written summaries, in the form of report cards, several times a year. In many households, the arrival of the report card is a significant event.

Teacher-written assessments of their own students are, then, an accepted part of the education system. Much more controversial, however, is the concept of standardized testing, where almost all students in the province work on the same material, which is then marked outside the local school to a uniform standard. Beginning in the 1996–97 academic year, Ontario introduced just such a standardized assessment of all its elementary students. The testing is carried out in May, graded over the summer, and returned to schools, students, and parents in the fall. At first, tests were administered only to Grade 3 students; two years after its introduction, testing was extended to Grade 6 students.

Opinions vary about the usefulness of these external assessments. The controversy really heats ups, however, when the external test results are averaged across students at each school, made available to the public, then used as a measure of the school's effectiveness. Is such a school report card useful? Are schools that contain students with higher average test results actually better schools? If better student test results are *not* an appropriate measure of a school's success, what is? Is the Ontario provincewide elementary assessment program useful?

One important group in the elementary educational process, Ontario's elementary school teachers as represented by their federations — the Elementary Teachers' Federation of Ontario (ETFO) and the Ontario English Catholic Teachers' Association (OECTA) — is among those who oppose standardized testing in its current form. Teachers particularly dislike the publication of results averaged over students at individual schools since such results are often used, implicitly or explicitly, to rank schools. ETFO argues that "[t]he best way to judge schools is by visiting them and looking for evidence of learning and interest in learning," and claims that "average test scores are *more* indicative of the pooled characteristics of the students, such as socio-economic status" (ETFO 2001b, 1–2; emphasis added). In fact, slightly *less* than half the variation in schools' results can be associated with the pooled socio-economic characteristics of their students examined in this book. So a large part of the explanation for the variation must be, at least to some extent, the outcome of choices principals, teachers, boards, and parents have made for their schools.

OECTA comments that school-level results are "poor measures of school success," and argues that

> research has repeatedly demonstrated that the comparative advantages and challenges children face outside of school — like their parents' socio-economic status and level of education — have an enormous impact on school success. To judge schools' performance — even in a carefully delimited way — to judge only students' comparative performance on paper and pencil

> tests of reading, writing and mathematics at a certain stage of their schooling — on the basis of their raw scores is patently unfair. (OECTA 2002, 16.)

Everyone involved in the education system knows that students vary greatly according to the social and economic background of the households in which they live. In some schools, many students have well-educated parents. In others, there are many recent immigrants. Some schools are large, others are small. Some schools have many aboriginal students, others have none. Most people would agree that direct comparisons of the test results of schools with such disparate groups of students would be unfair to teachers, boards, students, and parents alike. The Fraser Institute's report cards (Cowley and Easton 2003, 2004) for Ontario elementary schools, as I show in Chapter 6, are in fact just such direct comparisons of test results between schools with disparate student populations. These report cards, in my view, are unfair and not very useful. But does this mean, as some argue, that a fair and useful comparison is impossible?

I argue that it is not — that a fair and useful ranking and comparison of Ontario's elementary schools is indeed possible. In this book, I attempt to outline such a methodology and answer critical questions about the role of standardized testing in Ontario's elementary education system.

Toward a Fair and Useful Comparison

A comparison should not look simply at absolute test results and suggest that a school containing students with a high average mark is a "good" school; rather, it should focus on the relative ranking of schools after taking into account variations in the characteristics of the communities in which the schools are located. In that way, one can compare schools whose students have similar socio-economic background, and identify schools that have had substantially above-average assessment results over a number of years. This would be a useful and powerful exercise.

Moreover, we know enough about the backgrounds of Ontario's students to undertake such an exercise. As I outline in Chapter 6, one can predict the average assessment result over a number of years of a particular school based on both a wide variety of known socio-economic factors concerning the school's student body and the estimated relationships between those factors and test results at all Ontario elementary schools. The relevant factors include:

- the type of test and the year in which it was administered;
- the percentage of students living in single-parent families;
- the percentage of students speaking an official language at home;
- the percentage of students coming from an aboriginal background;
- the percentage of recent immigrants among students;
- the percentage of students living in detached homes;
- the average household income of students' household units;
- the percentage of students that moved in the past year;
- the percentage of students that moved in the past five years;
- the unemployment rate in the school community;
- the percentage of adults in the school community that has not completed high school; and
- the percentage of adults in the school community that has some university education.

Once these factors have been taken into account for all schools, one can make a valid comparison between two schools with the same *predicted* assessment results but widely varying *actual* results. By way of analogy, consider a series of temperature readings on a specific day — say, May 21 — in locations all over North America. Given 50 years of data, the average of those past 50 observations in each location would be a good predictor of the expected temperature on May 21 in that location. Depending on the actual reading for that day, one can then state unequivocally that the temperature is either higher or lower than normal. In the same way, one can determine whether, over a period of, say, four years, a school's test scores are higher or lower than normal for that type of school. If the results are a *lot* higher, then we have identified a good school.

It should go without saying that the ability to identify relatively better schools is critical. Educational outcomes in elementary school are important: we want our children to be good readers and good writers, we want them to master the basic skills of mathematics, and we want them to move on to high school with these skills already in hand. We want this to happen for students from all social and economic backgrounds.

Once I had identified the schools that were clearly doing better than others of their type, I visited 13 of them and asked teachers, principals, and parents why they thought their school seemed to have a better learning environment than other schools with similar socio-economic characteristics.

Although much more work needs to be done to identify Ontario's successful and not-so-successful elementary schools and the reasons for their relative performance, I hope this book makes an important, if small, start on that project. If we can identify schools where good practices are making a difference, we can, in turn, identify those good practices. Just as important, we may be able to convince educators and parents that it is even possible to identify a good practice. I hope people who know a great deal more about elementary education than I do will follow in my footsteps and visit good schools to see how it is done.

A Summary of the Findings

In this section, I present a summary of my findings in a question-and-answer format that avoids some of the more technical details and, I hope, allows the reader to absorb the main conclusions of the study easily.

Question 1: Is it true that school assessment results from 1997 to 2001 show an improvement over time in the learning outcomes of elementary school students in Ontario?

It will frustrate readers to discover that this important question cannot be answered directly using the data in this project. The reason is that the aggregate level of achievement on any test depends on students' ability in a given year to master the material, the consistency of the test instrument over time, and the grading of the test material. A higher score could be the result of better learning by students, an easier test, or more relaxed grading of the same test. The available data simply cannot distinguish among these three possibilities.

Question 2: That is frustrating! Can nothing be said about whether Ontario students have actually improved over the four years of tests?

I can state only that, if the tests were designed correctly, were equally difficult in all years, and were graded consistently, then student achievement levels increased considerably between the 1998–99 and 2001–02 assessments. The goal of the current Liberal government of Ontario is for 75 percent of students to be able to meet the provincial standard by the 2007–08 academic year, but such a lofty goal is meaningful only if the tests remain equally difficult and are graded consistently over that time. It will be difficult to assess if either condition holds.

Question 3: That is hardly the most helpful answer! Doesn't any of the statistical evidence I am about to read tell me whether the assessment process was designed correctly and has been consistent over time?

Unfortunately, the study contains no direct evidence with which to answer this question, only shreds of indirect evidence. In my view, that indirect evidence broadly supports the conclusion that the tests were designed correctly. The fact that achievement results have risen slowly over time — as would be expected in a large and complex organization trying to effect improvement — is indirect evidence that the observed increase in results may actually reflect better performance by students.

There is also evidence that the rate at which students are, for one reason or another, exempted from the assessment process is fairly consistent from one school board to another, which suggests, though weakly, that the assessments are administered consistently. The Education Quality and Accountability Office (EQAO), the independent, arm's-length agency of the Ontario government that administers the tests, apparently puts a great deal of effort into designing and pre-testing the assessments, then grading them consistently over time. In the end, however, it is up to education researchers, not an economist, to determine if EQAO is succeeding in that task.

Question 4: Do the answers to questions 1, 2, and 3 mean that the assessment process is useless?

Absolutely not! The assessment process is not useless. Students and classroom teachers get results for individuals, classes, and schools, which identify the strengths and weaknesses of individual students and schools at a detailed level. This information can clearly be put to good use. At the school level, the aggregate results can be used to compare outcomes across schools in a fair way, as this book explores.

Question 5: When I compare the achievement results of different schools in my community, I notice that the best results are obtained by schools whose students come from households where the parents are more affluent and better educated than average. Is it possible to measure the family background of students at elementary schools so that we can make a fairer comparison?

The answer to this question is a resounding yes! In academic years 1999–00, 2000–01, and 2001–02, the Ontario Ministry of Education collected the postal codes of all students in the province's elementary schools. By linking the location of students' homes through their postal code to very small geographic units of the census, one can, at little cost, draw a complete and accurate picture of the family backgrounds of students at any elementary school in Ontario. Chapter 3 explains this process in detail.

Question 6: I have read that good assessment results occur only in schools whose students come from affluent, well-educated families. Is that true?

No, this is not true. There is certainly a strong relationship between a variety of social and economic variables that describe the school community and school achievement levels. But, and this is the most important finding of the study, *some schools with less affluent and less well educated parents have high achievement results, and some schools with affluent, well-educated parents have results that are not especially good given their school community.* This finding has important policy implications. Since family background evidently is not the only determinant of school achievement, we need to discover why schools with similar socio-economic characteristics perform differently on the tests year after year.

Question 7: What socio-economic factors play a role in predicting which schools will have high achievement levels?

Higher income levels and better-educated parents certainly raise achievement levels, but other variables — including language, unemployment, mobility, housing status, immigration, aboriginal status, and lone parenthood — matter as well.

Question 8: Can schools be ranked according to outcomes in a fair way?

Yes. As I show in Chapter 6, it is possible to create a score that identifies whether schools with similar socio-economic characteristics are performing better or worse than other schools where assessment results should be similar.

Question 9: What should I do when I discover a "good" school?

My basic answer is that you should visit the school and try to find out how it differs from similar schools. What does that school do that allows its students learn more? As a small step toward answering that question, I undertook short visits to 13 schools (in six school districts) that substantially outperform equivalent schools (see Chapter 7).

Question 10: What did you learn about "good" schools?

That is a difficult question to answer simply, and I refer the reader to Chapter 7 for details. However, the single most important lesson I learned was that, in a successful school, teachers in the primary division (kindergarten to Grade 3) work together as a team to approach the Grade 3 test. Teachers in the junior division (Grades 4 to 6) also work together as a team to approach the Grade 6 test, but to a lesser extent. I also learned that longer visits to such schools are certainly warranted.

Question 11: Is the EQAO assessment process good for Ontario schools?

The answer to this question has to be formed from a subjective assessment of both the evidence and the benefits and costs of the EQAO process. Although the monetary costs associated with the process are relatively small, the indirect costs are much higher. There is a wide range of benefits from the process, but these are difficult or impossible to quantify. My personal conclusion, which I explain in Chapter 8, is that the benefits outweigh the costs, and I would choose to continue with the assessment process.

Question 12: Could the EQAO assessment process be improved?

Yes. Chapter 8 contains my recommendations for improvement, but the three most important are as follows. First, since the tests cover accumulated knowledge from three (or more) years of material, they should be renamed the "Primary Assessment" and the "Junior Assessment." Students' results are the primary responsibility, not of the Grade 3 or Grade 6 teacher, but of the school's divisions and of the school as a whole. Renaming the tests would make it clear that they are a school project, not a grade project, and would change the dynamic around the assessment process in a useful way. Successful schools, in my view, have already made this change.

Second, the test data should be presented to the public in a more meaningful way that takes into account the nature of the assessment process. For example, there is no point in discussing results where the key variable is *not* the percentage of all students at a school that achieved a satisfactory level. EQAO's "Method 2" measures the percentage of students that actually wrote the test and that succeeded. Since a school could claim better results simply by exempting from the test more students who would not achieve a good score, Method 2 results are not useful. Moreover, the results could be presented in a way that de-emphasizes changes from year to year and instead emphasizes multiyear results at a school. It would also be very useful to have results that adjust for the socio-economic characteristics of the school community. Indeed, that is the exercise I undertake in this book.

Third, the assessment should take place as late as possible in the school year, as participants in the process clearly prefer.

A Guide to the Rest of the Study

In the remainder of this study, I explain how a long and careful look at the assessment data leads to the answers I gave above. Chapter 2 provides the background of the elementary school assessment process, including some history, reactions to assessments, and estimates of their monetary cost. Chapter 3 describes in detail the process used to create a socio-economic description of each elementary school. Then, Chapter 4 presents estimates of the association between a school community's socio-economic variables and the school's level of achievement on the assessments.

In Chapter 5, I offer a brief discussion of the reasons students are exempted from the assessment process, which provides indirect evidence of the consistency with which that process is applied across Ontario's elementary schools.

Chapter 6 further explains how I created relative rankings for equivalent schools. Suppose the students of two different schools come from communities where the mix of socio-economic factors leads one to expect the two schools to have similar assessment scores. If their scores differ over a number of years — if one school regularly outperforms the other — then it is possible to devise relative rankings for the two schools. I rank the schools in one school board using this relative system and compare my results with the ranking the Fraser Institute has devised for the same schools. I show that any ranking — such as the Fraser Institute's — that depends mostly on the absolute level of assessment results and fails to take into account variations in the schools' socio-economic context is, in large part, a ranking of the schools' socio-economic status, and thus is of limited use.

In Chapter 7, I report on the visits I made to 13 "good" schools in 6 school boards — schools that perform relatively better, in terms of their absolute assessment results, than similar schools in the province. I visited schools that were expected to perform well and that did even better than expected, as well as schools that were not expected to do well but that exceeded expectations. I spoke to principals, teachers, and parents at these outstanding schools and simply report what they told me.

Finally, in Chapter 8, I offer my own opinions and conclusions about Ontario's assessment process for elementary schools, and I close with a set of recommendations that I believe would improve the process. These recommendations are directed mainly to EQAO and to the local schools, since I am much less confident of my ability to make recommendations at the school board level. Instead, I pose a number of questions for school boards that arose from my research and my visits to schools.

This book has a variety of audiences in mind. Some readers may be interested in the use of postal codes and census data in Chapter 3. Others may want to know a lot about the statistical properties of assessment data I analyze in Chapter 4. Some may care only about exemptions, discussed in Chapter 5, while others will want to go immediately to the school-ranking process in Chapter 6. Still others may want to read only about the visits to "good" schools, described in Chapter 7. And some may be interested only in my conclusions from this research, found in Chapter 8.

I hope these various audiences can find the material that is of most interest to them; the summaries that appear at the end of Chapters 2 through 7 should aid their search.

Finally, for readers who wish to examine my data in detail, the database and individual school data are available from the C.D. Howe Institute's web site at: www.cdhowe.org.

2

Ontario's School Assessments: Background Issues

In Ontario, assessments of elementary school students are undertaken by the Education Quality and Accountability Office (EQAO), established in 1996 as an independent, arm's-length agency of the provincial government. Its goal is, in the words of its first annual report, to "assure greater accountability and contribute to the enhancement of the quality of publicly funded education in Ontario."

Ontario is hardly alone in carrying out a jurisdiction-wide assessment of elementary school students — indeed, the practice is common throughout Canada and the United States. British Columbia, Alberta, and Quebec test their elementary school students for literacy and numeracy, and present the results publicly by school and by board, while New Brunswick and Nova Scotia have announced plans for a similar type of assessment. Saskatchewan tests a random sample of its students but does not reveal the results by school. Manitoba tests all Grade 6 students on three different subjects on a rotating basis and makes the results public by school. Newfoundland and Labrador uses the Canadian Test of Basic Skills in Grades 4 and 7 but does not report the results by school. Prince

Edward Island is the only province that does not assess its elementary school students.[1]

In the United States, the goal of the federal *No Child Left Behind Act of 2002* is for every child to be able to read by Grade 3. To measure progress toward this goal, the 50 states (and Puerto Rico) are encouraged, through the promise of federal funding, to develop appropriate assessments of academic achievement based on statewide curricula. As of February 2003, 21 states had qualified to receive such funding, according to the U.S. Department of Education.

Assessment data are also available for most schools in the United Kingdom and Australia.

EQAO's Activities

Although EQAO is best known in Ontario for its provincewide elementary and secondary school assessments,[2] the agency actually takes part in a broad range of important multijurisdictional assessment activities. Some, such as the School Achievement Indicators Project, are part of a nationwide assessment project. Others, such as the Third International Mathematics and Science Study or the Programme for International Student Assessments are, as their names suggest, international in scope.

In the first two years of its existence, EQAO administered a provincewide assessment of the reading, writing, and mathematics skills of the approximately 130,000 Grade 3 students in publicly funded elementary schools in Ontario. In the third year, the assessment program was expanded to include Grade 6 students, approximately doubling the number tested. All students in those grades are assessed except for a few who are exempted, with the agreement of their parents, if it is felt that their participation would be harmful.

---

[1] For a list of assessments by province, see the Canadian Teachers' Federation web site: www.ctf-fce.ca.

[2] EQAO, which at first tested only elementary school students, has now extended its activities to the secondary school level. The provincewide Grade 10 literacy test was first administered in the 1998-99 academic year, and the Grade 9 mathematics test became provincewide the following year. Recently announced changes reduce the length of the assessments to a total of six hours, presumably over three days (EQAO 2004).

Both the Grade 3 and Grade 6 assessments have a fairly clear cycle and format over the year. In the spring, all students in those grades at publicly funded schools work through a unit of curriculum for five days for about three hours per day.[3] In addition to test booklets on reading, writing, and mathematics, a multiple-choice component helps to calibrate the difficulty of the assessment from year to year.[4] Detailed procedures are laid out for administering the assessments, including where and when students may obtain help from their teacher. For example, the teacher may not help students read during the reading assessment, but during the mathematics assessment, the teacher is permitted to read the questions *verbatim*. For students who have difficulty writing, the teacher may write a mathematics answer *verbatim*. EQAO also randomly checks the assessment process by having its inspectors visit schools to ensure that the tests are being undertaken in an equivalent way across schools. The agency reports that it observed about 10 percent of classrooms during the 2001–02 assessment period.

EQAO has involved teachers, principals, and consultants in the creation of the assessment process from the beginning. At the time of the first assessment, however, there was considerable controversy about whether it was closely related to the curriculum; since then, the material in the assessments has been based on curriculum expectations laid out by the Ministry of Education. Material for the Grade 3 assessment is pre-tested on a sample of Grade 4 students while the Grade 6 assessment is pre-tested on a sample of Grade 7 students. The initial sample size was about 2,600 students; in more recent years, the field testing has involved about 600 students and 30 classrooms each year; teachers keep notes, and the assessments are subsequently revised as necessary.

---

[3] This procedure is designed to simulate the type of activity that would "normally" take place in a classroom rather than to create an atmosphere similar to a university or high school examination. Opinions differ on whether this goal is achieved. In the first year of the assessment, the curriculum unit was ten days, but it was shortened to five days in subsequent years in response to complaints about its length.

[4] For sample questions and answers, as well as a wealth of other information about the assessment process, see EQAO's web site at: www.eqao.com.

Test materials are not marked at the school; instead, they are collected and marked centrally by teams of graders during the following summer. Consistency in grading is an important goal — to that end, a significant number of already-marked test booklets are, on a random basis, marked again by a different grader. In this way, the quality of marking by individuals is checked and sometimes rejected.

Reporting the Assessment Results

EQAO reports the assessment results for the province as a whole, for each board, and for each school except where data are occasionally suppressed for privacy reasons in very small schools. In the first assessment year, most boards did not release school-level results, but public pressure quickly caused a change of heart, and results are now issued for the majority of schools in the province.

Students who have completed the assessment are placed in one of four levels. Those in Levels 1 and 2 "demonstrate and communicate limited to some knowledge, understanding and application of [reading, writing, and mathematics] skills," while those in Levels 3 and 4 "demonstrate and communicate extensive to in-depth knowledge, understanding and application of [reading, writing, and mathematics] skills" (EQAO 1999, 10). When the assessment process began, emphasis was placed on the percentage of students at Levels 2, 3, and 4 — that is, on students who showed at least some knowledge of the material. Now, however, the stress is on the percentage of students at Levels 3 and 4, which, over time, have come to be regarded essentially as the levels at which students meet or exceed curriculum expectations, while students at Levels 1 and 2 are viewed as failing to meet expectations.

Using the Assessment Results

How are all these data put to use? The public focuses on overall scores in reading, writing, and mathematics at the board and school levels and for the province as a whole. However, these scores are aggregations of much more detailed grades that schools (and parents) receive, and much of the analysis is hidden from public view and used for internal planning purposes. For example, each assessment area consists of "strands," or identifiable skills, and students who exhibit deficiencies in these areas are made known to their school, which is then expected to adjust its teaching accordingly. A similar analysis takes place at the board level. In addition, schools and boards file responses to EQAO assessments.

The Data's Public Face

The public face of the assessment data has two aspects. One is the use of the data to determine whether students in the system have improved over time; the other is the use of the data to "rank" schools. There are, however, difficulties with both uses.

One of the problems with using the data to measure improvement over time is that the first assessment, which occurred in the 1996–97 academic year, cannot be compared validly with subsequent, modified assessments. EQAO claims, however, that at least some of the results from the later assessments can be used to monitor overall performance at the provincial level:

> The overall reading and mathematics scores from the 1997–98 and 1998–99 assessments can be compared from year to year. The reading and mathematics components each contain a set of multiple-choice questions which overlaps from year to year. EQAO is in the process of developing a mechanism to ensure compatibility of the overall writing scores. (EQAO 1999, 8.)

EQAO repeats this claim, with less equivocation, in its 2001–02 report: "Overall levels of achievement are comparable from year to year" (EQAO 2002, 8).

From an outsider's point of view, generating exactly equivalent assessments graded in exactly equivalent ways, even on a very large sample, seems a daunting task.[5] Indeed, as I discuss in Chapter 4, the assessment results appear to have fluctuated over the years, and it is not a straightforward matter to interpret the results in the light of such fluctuations. Fortunately, much of the analysis of the assessment data is valid without the need to assume that the instrument is exactly comparable from year to year (again, see Chapter 4).

It is the second public face of the assessment data — the use of the data to compare schools and, to a lesser extent, school boards — that excites the most comment and even controversy. Data on individual schools have been made public since the second assessment year, 1997–98. These results typically are seized upon by local news outlets, which publish tables of rankings, with their implicit suggestion of the existence of "good" and "bad" schools. Parents naturally look for the ranking of the school their children attend, and the results are discussed at school councils.

Accounting for the Variance in Schools' Results

It is no secret that the schools with the higher assessment results are in neighbourhoods with more favourable socio-economic characteristics. Indeed, EQAO's first annual report seems to have acknowledged this outcome in its recommendation that "a wide range of data, including social-economic information and specific regional factors, [should] accompany future assessment results" (EQAO 1997, 70). Yet, until now, there has been no detailed, comprehensive analysis of the socio-economic context of school performance.

Two studies of Ontario's elementary assessment results do take the schools' context into account, but in a limited way. One study

---

[5] The Provincial Auditor of Ontario expresses similar doubts: "At the time of our audit, the EQAO's procedures for designing and reporting results on its large-scale tests had not been subjected to an independent examination by experts in the field of psychometrics....It is expected that the review will be completed by March 2004" (Provincial Auditor of Ontario 2003, 134).

(reported in Tremblay, Ross, and Berthelot 2001, 2002) examines the results of the first assessment, in 1996–97, using census data for a broad, 1.6 kilometre radius around the school location.[6] Using such a crude measure, the authors identify income and the education level of the population of adults as important factors associated with higher school scores. They also find that urban schools outperform rural schools, that children of immigrants actually do better than the native born, and that school size has little effect on performance. As we will see later in this book, these findings parallel my own.

Harris and Mercier (2000) also use census data to examine the socio-economic context of the 1996–97 assessment results for a sample of schools in Hamilton and metropolitan Toronto. Again, however, their unit of measurement is very broad — namely, the relatively large geographic area of the census tract — and again they find that the income and educational background of parents are important contributors to higher school results.

By far the largest public exercises undertaken with the assessment data are the report cards published by the Fraser Institute (Cowley and Easton 2003, 2004), which cover many of Ontario's elementary schools. However, the explicit numerical rankings in these report cards are based on the schools' absolute achievement levels, and essentially ignore the socio-economic characteristics of the communities in which they are located.

In this book, I greatly expand the coverage of these studies by examining several years of data for all elementary schools in the province. I use both the 1996 and 2001 censuses to measure precisely the socio-economic characteristics of the communities in which the schools are located by linking individual students' postal codes to the smallest geographic unit of the census in which the students live.[7] Although the Fraser Institute's report cards imply that a school's success depends mostly on the socio-economic background of its students, what does "mostly" mean? My goal is to provide a quantitative answer to this absolutely critical question. It turns out that useful information can be found both in rankings (such as the Fraser Institute's) that do not depend on socio-economic context and in those (such as the ones I develop) that compare schools with similar socio-economic backgrounds. I discuss and compare the methodologies of both types of rankings in Chapter 6 — I leave it to the reader to decide which is the more useful.

---

[6] The study does make use of some EQAO data collected at the level of the student — such as the student's sex, whether or not the student speaks English as a second language, is in French immersion, has many books at home, and has a computer at home. Some classroom-level data — such as the teacher's experience, class size, and whether or not there is a computer in the class — are also used.

Similar Studies in Other Jurisdictions

To restate, if assessment results are to be used in a meaningful way, they need to be placed in context to see if, given students from similar backgrounds, they vary significantly across schools. This is the task I have set for myself in this book with respect to Ontario schools, but it has also been undertaken in other jurisdictions using two different methodologies.

The ideal methodology should produce a standardized test result from the *same* student at *different points in time*. With even better data, both the teacher and the school would also be identified, which would allow the researcher to determine if different teachers or different schools play a role in changes in the results. Although these kinds of data are rarely available to the researcher, some such studies have been undertaken.

Using just such data for 10,000 students and 300 teachers over a ten-year period in two New Jersey school districts, Rockoff (2004) concludes that teachers have a strong impact on standardized test results. In another study, Sanders and Horn (1994) use data linking students to individual teachers in the Tennessee Value-Added Assessment System reach a similar conclusion: teaching matters!

---

[7] My work also builds on Dickson (2001), a pilot study of assessment data for the 1996–97 through 1998–99 academic years for public elementary schools in Waterloo County. Dickson's census characteristics of school communities were derived by matching the smallest census geographic units to school boundary maps provided by the school board.

In the absence of data linking students to specific teachers, however, there are two other possible ways to measure school quality. One way is to use data sets that contain multiple observations on the same student at the same school, so that, even without a link to specific teachers, it is possible to ask whether the progress of students at one school is more rapid than that of students at another school. Several recent U.K. studies have done just that (see, for example, Wilson 2004). Although the United Kingdom has had standardized tests at the secondary level for decades, only recently have school results been published in what are called "league tables," and students at elementary schools are now also tracked. In December 2003, students in 14,644 primary schools who had been tested at age 7 and again at age 11 were tracked and the progress measured of those who had remained in the same school.[8] Not surprisingly, schools with a large proportion of students whose parents are affluent and well educated inevitably perform better than schools with students whose parents are poorer and less well educated. While such comparisons are not perfect, since students are mobile and schools' retention rates vary widely, they are clearly better than ranking schools based simply on the actual assessment results.

The school system in Dallas, Texas, also provides data on the performance of individual students at the same school as well as a number of the students' socio-economic characteristics. With such data, it is possible to calculate the value added by the school for these students and to compare across schools.[9] Critics of this method of analysis argue that the target is, or should be, an absolute standard of primary literacy and that large improvements starting from a very low level may not be an appropriate measure of performance — that is, although a school may be successful in increasing the literacy of its students, if the starting point is low enough the "improved" student may still be illiterate.

---

[8] For a useful web site describing various issues surrounding the primary value-added tables, see www.education.guardian.co.uk/primaryeducation. See also the U.K. Department of Education web site: www.dfes.gov.uk/performancetables.

[9] The methodology used to construct value-added rankings by school is described in Clotfelter and Ladd (1996).

Another way to measure school quality is to look at the aggregate data on student performance at the level of the school in a given year. This is the most widely available type of data and, indeed, it is the form the Ontario data take. By far the largest such exercise so far looks at all schools in Texas over academic years 1999–2000, 2000–01, and 2002–03 (National Center for Policy Analysis 2003). Using data on various demographic and social characteristics of individual students, the study predicts a score for each Grade 6 student, then aggregates the difference between actual and predicted scores by school to measure if a school performed better or worse than predicted.[10] Another study, of the schools in a Maryland suburb of Washington, D.C., also makes a serious attempt to take into account variations in students' socio-economic backgrounds.[11]

No matter how schools are compared, however — whether on a value-added basis using individual student data or by looking at students' socio-economic backgrounds — there is still considerable controversy about the usefulness of the exercise. Rowe (1999), whose cautions I take into account in this book, points out that, by definition, all measures are based on the school's past, rather than current, performance. He stresses the need to look at many years of data, though, of course, that means having data that stretch back into the past. Rowe also correctly points out that there is substantial statistical uncertainty in any value-added or relative school ranking — in other words, one has to be sure that one's methodology identifies schools that are, in the statistical sense, *a lot better* than other schools. My methodology accomplishes this important goal.

---

[10] For a description of the methodology used in the Texas study, see web site: www.mykidseducation.org.

[11] See web site: www.mcps.k12.md.us.

Political and Institutional Reactions to the Ontario School Assessment Process

Standardized provincewide assessment of Ontario's elementary students remains controversial, seven years after its introduction. In their fall 2003 election platforms, two of Ontario's three major political parties expressed clear support for testing in something like its present form.[12] For its part, the New Democratic Party supported continued testing but wanted changes — in particular, that the raw test results should no longer be made public in such a form as to be able to compare students, teachers, schools, or administrators. The NDP's position was somewhat ironic considering the enthusiasm with which Bob Rae's NDP government had greeted the Royal Commission on Learning's 1995 suggestion to establish EQAO, by announcing funding of $25 million for the project. After the New Democrats' defeat, the new Progressive Conservative government reduced the amount to $14 million.

Although the Liberals' 2003 election platform was silent on EQAO, a key platform promise was to have 75 percent of students "meet or exceed the provincial standard on provincewide tests" in a Liberal government's first mandate. What this implies for the future for EQAO assessments, including the public provision of school-level data, is not entirely clear, since such a promise could be met without the release of any school-by-school results.

In contrast to the apparent broad political support for testing, Ontario's two federations of elementary school teachers — the Elementary Teachers' Federation of Ontario (ETFO) and the Ontario English Catholic Teachers' Association (OECTA) — remain opposed to standardized testing in its present form (see ETFO 2001a, 2001b, 2003; OECTA 2002). Among their objections are that

---

[12] At the time of the election campaign, the parties' platforms were available at the following web sites: www.ontarioliberal.com/en/platform; www.ontariopc.com/TheRoadAhead; and www.publicpower.ca (for the Ontario New Democratic Party).

- students receive their results only in the fall, long after the spring assessment;
- the assessment procedure and format are disruptive to the normal flow of the classroom;
- the process is stressful to students;
- the multiple-choice portion of the assessment is outside the experience of a Grade 3 student;
- scores, particularly in small schools, fluctuate a great deal from year to year for purely random reasons; and
- the overall results at the school level are made available to the public without important background information about the context of the school.

Both federations argue that, if it is necessary to assess the overall quality of educational outcomes in Ontario, that task would be more effectively and more cheaply undertaken through the participation of random groups of Ontario students in multi-jurisdictional assessments. I address a number of these objections during the course of my analysis and discover that some have merit.

It is hard to ascertain the views of parents about standardized assessment. Using a rather nonscientific methodology, an electronic search through five southern Ontario newspapers over the period of EQAO testing yields a wide variety of opinions. There are very clear letters from a number of parents indicating their intention to withdraw their children from the assessment process, although there is no evidence of a widespread parental boycott of the process given that more than 95 percent of eligible children take part. It is reasonably clear that these vocal parents (and presumably many other parents as well) realize that the value to their child of the individual feedback that happens six months after the assessment is quite limited, and that the immediate feedback to a child from the classroom teacher is more meaningful.

School board officials appear to be cautiously supportive of the assessment process. Certainly, EQAO responded satisfactorily to boards' complaints about the length and difficulty of the 1997 assessment. Jim Daly, of the Hamilton-Wentworth Roman Catholic

Separate School Board, was quoted as saying "This [EQAO assessment] is much more reliable and constructive in enabling boards to improve because the tests the EQAO are developing are based on our curriculum. International tests are not" (*Hamilton Spectator*, November 1, 1997). The Ontario Catholic School Trustees Association, for its part, asks only that "reporting mechanisms must be reasonable in number and size and produce understandable and useful information."[13] In April 2004, the Board of Directors of the Ontario Public School Board Association (OPSBA) passed a motion that "the OPSBA supports the continuation of a provincial testing program for students administered by an external and impartial third party agency in order to improve instruction and preserve accountability to parents and the public." And in a May 18, 2004, press release that followed the publication of the second Fraser Institute report card (Cowley and Easton 2004), OPBSA President Gerry Gershon said, "Boards have always used EQAO results to help our teachers and schools to develop strategies to improve student learning and achievement." At the same time, however, she was highly critical of the Fraser Institute study, charging that "[t]oday's published ranking…undermines valid evaluation and testing measures, discourages and demoralizes teachers and belittles the efforts of our students."

My search of southern Ontario newspapers revealed few comments on the assessment process from individual teachers, most teachers apparently preferring their federations to speak for them. I found only one letter to the editor that was even partly supportive of testing, with teacher John Bradley writing that "[o]ne should not…be quick to dismiss these test results as useless" (*Hamilton Spectator*, August 10, 2001). He said that assessments were useful in reminding teachers to cover the curriculum, in acting as a catalyst for change, and as providing useful data to which responses were possible, and he argued that test results were a useful picture of a school at a point in time.

---

[13] See the association's web site: www.ocsta.on.ca/index.htm.

Clearly, however, many classroom teachers remain opposed to EQAO testing.[14] In another letter, the OETF's Diane Tomlinson reported that "[t]he pressure is unbelievable [on Grade 3 and Grade 6 teachers]. The teacher gets uptight. Her reputation is on the line" (*Hamilton Spectator*, May 13, 2000). On the other hand, and somewhat more positively, EFTO president Phyllis Benedict admitted that "[i]f there is something positive in EQAO testing, it is for the teachers who are lucky enough to be selected as markers" (*Hamilton Spectator*, July 13, 2000). This is a sentiment apparently widely shared among teachers, with 9,000 applications for 1,300 EQAO marking positions in 2000.

I do not pretend that my scan of southern Ontario newspapers offers anything like a complete picture of the responses of Ontario's schools, school boards, and parents to EQAO's elementary school testing process; rather, my intent was to suggest that those responses have been varied.

How Much Does the Assessment Process Cost?

One frequently voiced criticism of the assessment process is that testing costs a great deal of money that could be better spent in other ways — in other words, that the costs of the program outweigh the benefits.

Is this the case? Such a question is difficult to answer. The benefits and costs are complicated to measure, particularly since many of the costs are indirect. For example, school board officials, administrative staff, and teachers spend a considerable amount of their time planning and carrying out the evaluations — time that could be spent on other activities of perhaps more benefit to schools and to society at large. Students spend two and a half days writing the tests as well as some time preparing for them — again, time that could be spent on perhaps more useful schoolday activities. I am unaware of any studies that quantify such costs or consider the value of alternative uses of resources, and I make no such attempt to do so here. It is possible, however, to summarize estimates of the *direct* monetary costs of the assessment process.

---

[14] A 2003 COMPAS poll of Ontario teachers (COMPAS Inc. 2003), although not including direct questions about the value of assessments, seems to indicate that most teachers do not see student testing as improving students' learning or measuring students' success, neither do they see the assessment process as a good measure of schools' or teachers' success. It is not clear, however, whether these responses can be interpreted as opposition to assessments *per se*.

From fiscal year 1998/99, when the Public Accounts of Ontario first included a separate budget line for EQAO, to 2001/02, the agency's budget grew from $11,785,600 to $51,126,858. During that time, EQAO added the Grade 9 mathematics assessment and the Grade 10 literacy assessment to its tasks. EQAO also carries out the other multi-jurisdictional tests in which Ontario participates, although these are a relatively minor part of its duties.

Apart from the fixed costs associated with the testing process, any allocation of costs to specific assessments must be an estimate — and such estimates vary. The "most" precise might be that of the first two years of the program, when the tests were confined to Grade 3 students; thus, in 1997, the 138,000 Grade 3 assessments were estimated to cost $7 million, or $50.72 per assessment. An October 1998 Ontario government press release suggested that the 1998–99 Grade 3 and Grade 6 assessments, involving about 140,000 students, were expected to cost $6 million, or $42.00 per assessment. Still another estimate, from 2001, is $65.00 per assessment (*Hamilton Spectator*, December 5, 2001), while the Organization for Quality Education, a lobby group that wishes to replace the EQAO assessment with another type of test, estimates the cost as $95.00 per assessment, although it gives no indication of how it derived this number.[15] In any case, a sensible range of estimates seems to lie between $40 and $100 per assessment.

Is this a lot of money? Suppose the amount were as much as $100 per assessment. Since the assessment is given to students in

---

[15] See the organization's web site: www.oqe.org/7.html.

Grades 3 and 6 and there are approximately equal numbers of students in each grade from junior kindergarten to Grade 8 (ten grades), this is equivalent to $20.00 per student per grade from junior kindergarten to Grade 8. Scaled to the level of a classroom of 30 students, this represents $600 per year per room or, for a 300-student school, about $6,000.

Another way of estimating the cost is to take the amount that EQAO spends annually (about $50 million) and divide by the number of students in all grades in the Ontario education system (about 2 million). Thus, average EQAO spending per student per year is $25.00 or, for a 300-student school, $7,500. The $50 million EQAO spends each year is a lot of money, but it is spread over a lot of students and a lot of activities. Many of the teachers I interviewed believed that the monetary cost of the EQAO assessment program was equivalent to hiring an additional teacher at their school. This is clearly not the case.

Conclusion

In this chapter, I have tried to outline the extent of the controversy that surrounds the history, use, and cost of elementary school assessment in Ontario. The participants in the Ontario education process certainly do not agree on the usefulness of testing, and there is a paucity of research in the public domain using Ontario's results, which I hope to correct to some extent in this book.

Three questions arise from the discussion so far: Is the assessment process worth the cost? Is it possible to use the assessment results to discover which schools provide their students a better educational outcome? Can this comparison be done in a way that takes into account the socio-economic differences in the communities from which schools draw their students? I address the first question in Chapter 8, and offer my conclusion that the assessment process is worth the cost. To answer the second and third questions, I need to measure accurately the socio-economic characteristics of the schools' communities, a task I undertake in the next chapter.

Chapter Summary

Each spring, Ontario's Grade 3 and Grade 6 elementary school students take part in reading, writing, and mathematics assessments conducted by the Education Quality and Accountability Office (EQAO), an arm's-length agency of the provincial government established in 1996. The agency's goal is to "assure greater accountability and contribute to the enhancement of the quality of publicly funded education in Ontario." Ontario is among the numerous jurisdictions in Canada and the United States that carry out such tests of their elementary school students. EQAO reports the assessment results for the province as a whole, for each board, and for each school, and students are placed in one of four levels, with Levels 3 and 4 regarded as those at which students meet or exceed curriculum expectations.

There is considerable controversy about these test results, particularly when they are used to compare and rank individual schools, with the media typically seizing on the results to suggest that some schools are "good" while others are "bad." The best-known public examination of the assessment data is the annual report card published by the Fraser Institute. Its numerical rankings are based on schools' absolute achievement levels — not surprisingly, the schools with the higher assessment results are in neighbourhoods with more favourable socio-economic characteristics. Yet, if the test results are to be used in a meaningful way, they need to be placed in context to see whether there is significant variance among schools with students from similar socio-economic backgrounds.

While there appears to be broad political support for the EQAO assessment process, Ontario's elementary school teachers' federations are opposed to standardized testing in its present form, citing, among other objections, the length of time between the tests and the release of results, the disruptive nature of the assessment procedure and format, the stress on students, the inappropriateness of multiple-choice questions, especially for Grade 3 students, and the often random fluctuation of test scores, particularly in small schools. For their part, parents seem divided over the usefulness of standardized assessments, while school board officials appear to be cautiously in favour of them.

One of the most frequent criticisms of the assessment process is its expense. Although the benefits and costs are complicated to measure, particularly since many of the costs are indirect, testing appears to cost between $40 and $100 per student tested, or about $25 per elementary school student in Ontario. For a 300-student school, that amounts to about $6,000, which is far less than the cost of hiring even a single additional teacher at the school.

3 Describing an Elementary School's Community

In analyzing the results of any standardized test of elementary school students, the biggest problem is placing them in context. "Context" has two aspects. On the one hand, "context" means the year in which the test is written and graded. Despite efforts to create and grade assessments so that year-to-year comparisons are possible, tests may vary in difficulty and grading standards may vary over time, so that parents, teachers, and researchers simply do not know if the assessment instruments are comparable from one year to the next. Although this variance makes year-to-year comparisons of student achievement difficult, it is not a major concern in comparing schools, which is the issue of primary interest in this study. The year in which the assessment takes place is simply one of several context variables, which I explore in Chapter 4. In many ways, the year and type of test is the easiest variable to consider.

On the other hand, "context" refers to the socio-economic characteristics of the community from which a school draws its students. Everyone involved in the school assessment process recognizes that the socio-economic backgrounds of students are associated with how well they perform on the tests. But, as I noted in Chapter 2, some participants in the process then draw on this well-known association to argue that provincewide testing is thus a useless exercise. Does the evidence support such an argument?

To answer this question, one must first construct measures of the socio-economic characteristics of the school community, which is the task I undertake in this chapter. Once one takes into account the effects of students' socio-economic backgrounds on variations in their test results, it becomes possible to see if schools vary in systematic and interesting ways and to speculate on the reasons.

Measuring School Community Characteristics: A General Description

Ideally, one would like to be able to measure directly each individual student's socio-economic characteristics and associate them with the student-level assessment results. Such data are not collected, however, because of concerns about privacy. Instead, one is able to measure only the average characteristics of households in the community from which a school draws its students — which, for convenience, I call "school characteristics" or "school community characteristics" — and then only indirectly.

To determine those school community characteristics, one must turn to the census, which asks detailed questions of a 20 percent sample of Canadian households. The answers are reported as the average characteristics of persons or households in a variety of geographic units, including very small ones — for example, in the 1996 census, the smallest units, known as Enumeration Areas (EAs), contain between 125 and 650 dwellings; in the 2001 census, the smallest units, known as Dissemination Areas (DAs), are 400 to 700 persons. So, even though we do not know the socio-economic characteristics of individual students, the characteristics of the small census units in which students live can tell us much about them. And we can determine which small census units students live in if we know their postal codes.[1] Fortunately for the researcher, the

---

[1] Note that, since the geographic boundaries of a school's catchment area differ from the EA or DA in which particular students live, the use of postal codes...

Ontario Ministry of Education has compiled files, by postal code, of the number and sex of students for most elementary schools in the province for the academic years 1999–2000, 2000–01, and 2001–02.

To provide a sense of the scale of the school communities in question, Table 3.1 presents the number of schools as well as minimum, maximum, and average enrollment numbers in both the public and separate school systems as reported in Ontario's postal code files and the 1996 and 2001 censuses (note that the number of schools increased substantially over the three academic years). For the 2000–01 academic year, at least, the postal code files and census data cover about 95 percent of Ontario's elementary schools and students, as Table 3.2 shows.

What the Census and Postal Code Data Say about School Community Characteristics

The census data tell us a great deal about the socio-economic characteristics of school communities, including information about immigration, language, the number of aboriginals, education status, family status, various economic indicators, as well as mobility and housing. And, as noted above, the postal code data tell us about school enrollment and the sex of students. Let us take a closer look at these variables.

Immigration and Language Variables

Immigration and language are frequently used as important context variables for school communities. In some cases, for example, Ontario's Education Quality and Accountability Office (EQAO), which

---

Note 1 - cont'd.

...captures students who, for various reasons, cross school catchment boundaries. For example, a significant number of students in French immersion programs live outside the catchment areas of the schools they attend. Note also that the postal codes used in this chapter are those of students, not of the school they attend.

Table 3.1: *Coverage of Postal Code and Census Data for the Ontario Elementary School System*

Year of Postal Code Data	Number of Schools		Minimum Enrollment		Maximum Enrollment		Average Enrollment	
	Public	Separate	Public	Separate	Public	Separate	Public	Separate
1999–2000	2,185	1,223	0	7	2,119	1,887	355.9	345.8
with 1996 census	2,185	1,117	0	3	2,119	1,783	355.9	330.0
with 2001 census	2,185	1,123	0	3	2,119	1,870	355.9	343.0
2000–01	2,460	1,292	1	2	1,333	1,810	368.1	338.4
with 1996 census	2,459	1,291	1	2	1,322	1,669	351.8	316.9
with 2001 census	2,460	1,292	1	2	1,333	1,785	368.1	338.3
2001–02	2,563	1,370	4	3	1,429	1,677	350.4	319.4
with 1996 census	2,563	1,370	4	3	1,428	1,635	342.8	309.6
with 2001 census	2,563	1,370	4	3	1,429	1,675	350.4	319.4

Note: A few schools have an enrollment of zero; these appear to be administrative units to deal with certain special-needs students that are not active in a particular year. Without students, they accordingly play no role in the subsequent analysis.

Table 3.2: ***Participation in Postal Code Collection, academic year 2000–01***

	Number of Schools		Number of Elementary Students	
	Public	Separate	Public	Separate
In provincial data	2,573	1,390	958,489	476,256
In postal code sample	2,460 (95.6%)	1,333 (95.8%)	905,526 (94.5%)	437,213 (91.8%)
In 1996 census	2,459 (95.6%)	1,322 (95.1%)	865,076 (90.2%)	409,130 (85.9%)
In 2001 census	2,460 (95.6%)	1,333 (95.8%)	905,526 (94.5%)	437,083 (91.7%)

Sources: Ontario Ministry of Education web site: www.edu.gov.on.ca/eng/general/elemsec/quickfacts/2000-01/index.html01; author's calculations from Ministry of Education postal code files of elementary school students.

conducts the standardized tests, has presented separate results for English-as-a-second-language (ESL) students and non-ESL students.

There are two immigration variables: the percentage of the school community's population that has arrived in Canada in the past five years, and the percentage of the school community that was born outside Canada, the United Kingdom, and the United States.[2] Recent immigrants account for less than 5 percent of the school community population of many schools, but in a few they are more than 20 percent. Persons born outside Canada, the United Kingdom, and the United States are much more evenly spread throughout the school system than are recent immigrants (see Table 3.3). Figure 3.1 shows the distribution of the two immigration variables across school communities.

Language is also represented by two variables: the percentage of the school community whose mother tongue is either English or French, and the percentage that uses either official language as the

---

[2] The recent immigrants, those who have arrived during the past five years, are not broken down by birthplace or country of origin.

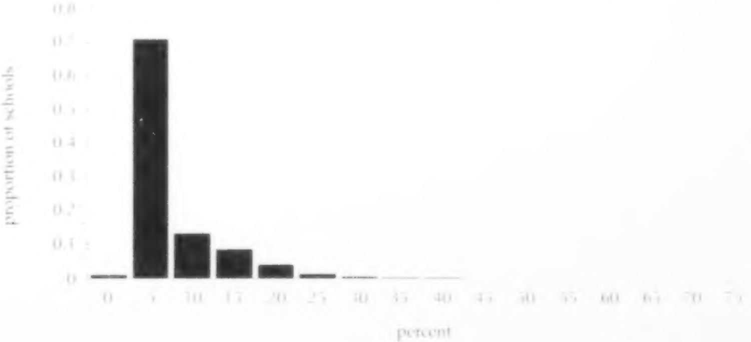
Table 3.3: *Descriptive Statistics for Schools in the Postal Code Sample*

Variable	Mean	Standard Deviation	Minimum	Maximum	Number of Schools with Extreme Values
Recent immigrants (in Canada < 5 years) (%)	4.3	5.6	0	43.9	2,908 < 5% 89 > 20%
Born outside Canada, US, or UK (%)	19.0	17.1	0	76.8	754 < 5% 932 > 30%
English or French as home language (%)	91.4	9.6	44.6	100	2,855 > 90% 6 < 50%
English or French as mother tongue (%)	79.6	16.2	23.9	100	1,424 > 90% 337 < 50%
Aboriginal status (%)	1.72	3.1	0	52.3	87 > 10%
Adults without secondary school diploma (%)	45.0	9.7	13.2	76.5	10 > 70% 0 < 10%
Adults with some university education (%)	22.0	11.2	0	75.1	318 > 40% 340 < 10%
Children in single-parent families (%)	21.9	8.5	0	63.6	137 < 10% 719 > 30%
Unemployment rate (%)	7.9	3.1	0	34.9	112 > 15%
Unemployment rate for persons with children (%)	6.5	3.5	0	39.6	104 > 15%
Average household income, 2+ person household (1995 $)	62,017	18,771	0	264,028	119 > 100,00
Moved in the past year (%)	13.5	4.1	2.3	51.9	15 < 5% 286 > 20%
Moved in past 5 years (%)	*41.1*	*9.6*	*10.6*	*87.3*	*0 < 5%* *2,137 > 40%*
Living in detached home (%)	65.6	22.9	0.32	100	312 < 25% 104 > 95%
Total enrollment	336	171	1	1625	228 < 100 677 > 500
Proportion of boys	0.498	0.04	0	1	52 > 60 66 < 40

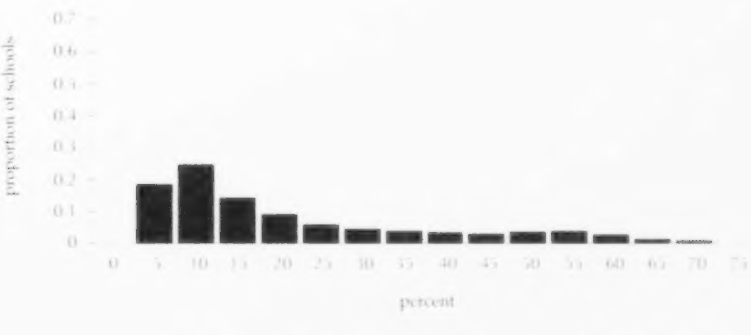
Source: Author's averages of postal code data from Ontario Ministry of Education for academic years 1999-2000, 2001-01, and 2001-02, and data from 1996 and 2001 censuses.

Figure 3.1: ***Immigrants in Ontario Elementary School Communities***

Panel A: Percentage of the school community that immigrated in the past five years



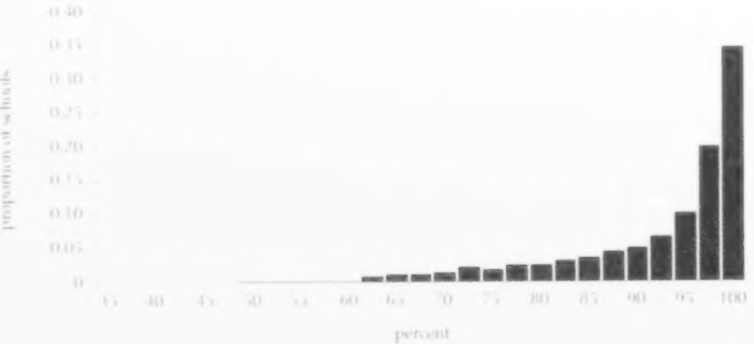
Panel B: Percentage of the school community that was born outside Canada, the United States, and the United Kingdom



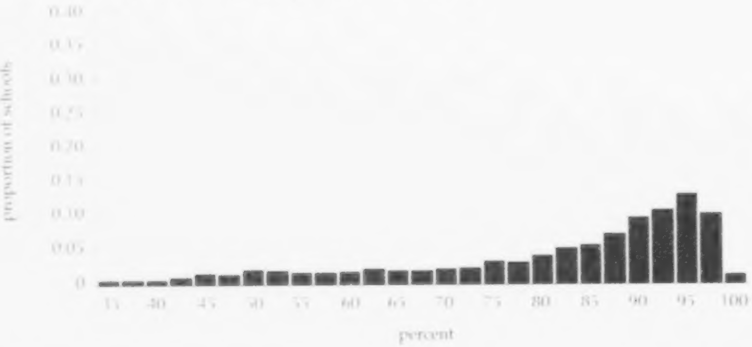
main language at home. On average, 90 percent of the school community population speaks an official language as the main language at home and 80 percent lists one of the official languages as the mother tongue (Table 3.3). In the vast majority of schools, more than 90 percent of students come from a language background that includes either English or French. Only in six school communities do fewer than half the students use either English or French as the main language at home. Figure 3.2 shows the distribution of the two language variables.

Figure 3.2: ***Language in Ontario Elementary School Communities***

Panel A: Percentage of the school community that speaks an official language at home



Panel B: Percentage of the school community that speaks an official language as its mother tongue

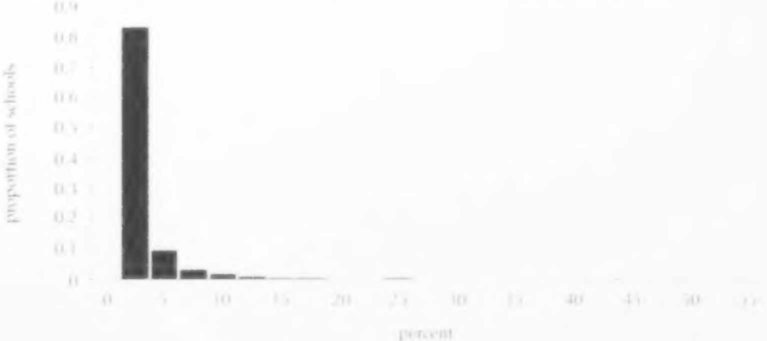


The Aboriginal Variable

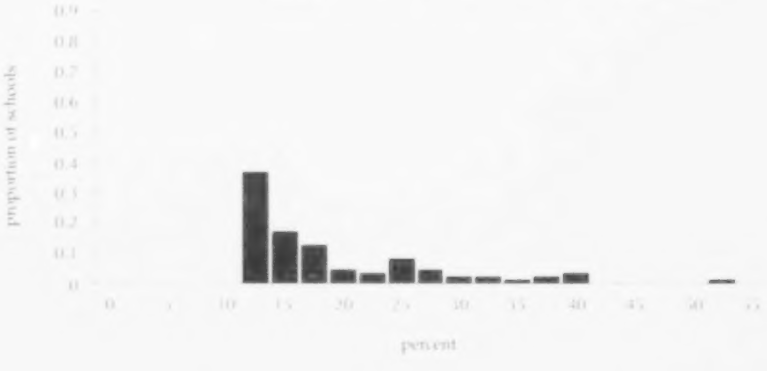
The aboriginal variable is simply the percentage of the population that claims to belong to any of the categories of aboriginals in the census. Figure 3.3 presents two pictures of the aboriginal population in Ontario school communities. Panel A shows the distribution of aboriginals across all elementary schools, the vast majority of which have a negligible aboriginal population. Only 87 schools in the entire province draw from communities that contain a population

Figure 3.3: ***Aboriginals in Ontario Elementary School Communities***

Panel A: Percentage of the school community with an aboriginal background



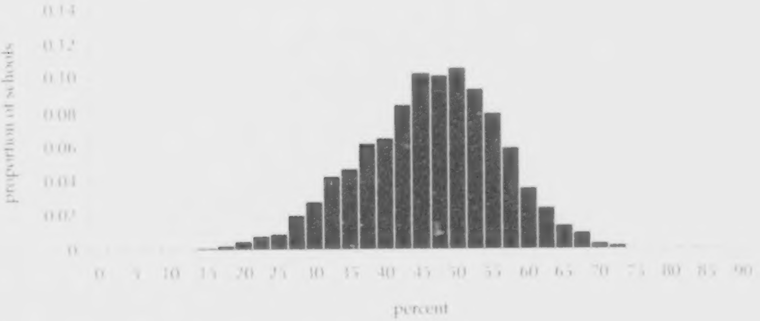
Panel B: Percentage of students with an aboriginal background in school communities with a population more than 10 percent aboriginal



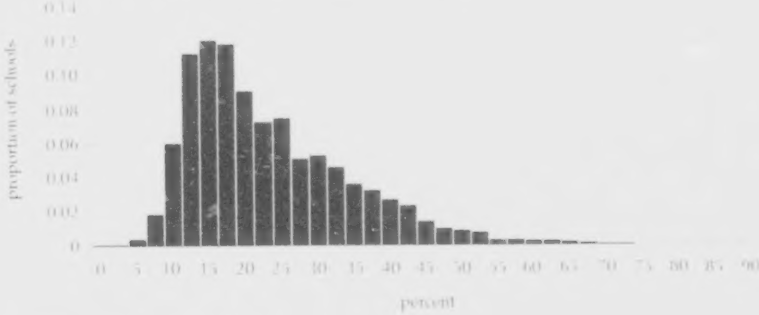
that ranges from 10 percent to 75 percent aboriginal, with the average being about 20 percent (see Table 3.3). Panel B of Figure 3.3 shows the distribution of the aboriginal variable in the remaining school communities — that is, those communities with a population that is more than 10 percent aboriginal. In summary, the proportion of aboriginals in Ontario school communities is very small for the vast majority of schools, but it is an important characteristic of a small number of schools.

Figure 3.4: ***Education Levels of Adults in Ontario Elementary School Communities***

Panel A: Percentage of the school community over age 15 without a high school diploma



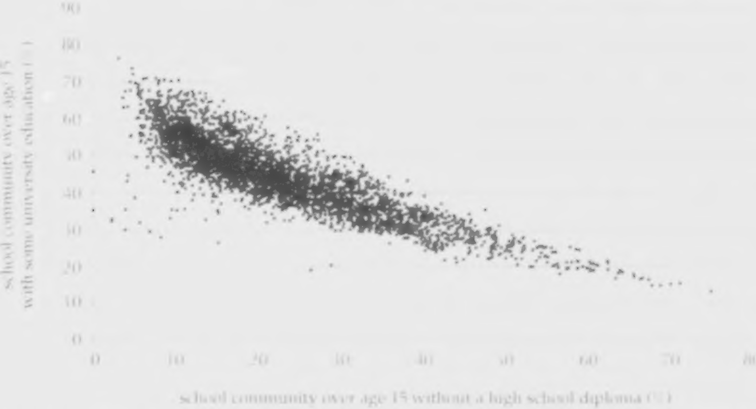
Panel B: Percentage of the school community over age 15 with some university education



Education Variables

Schools vary widely in the educational background of the adults in their community. We know this because of the detailed information in the census on the educational status of people over age 15. For our purposes, it is useful to look at two education variables that represent extremes in educational status: the percentage of people over age 15 in the school community without a high school diploma (Figure 3.4, panel A) and the percentage with at least some university education (panel B). In census areas where elementary

Figure 3.5: ***The Relationship between the Education Variables***



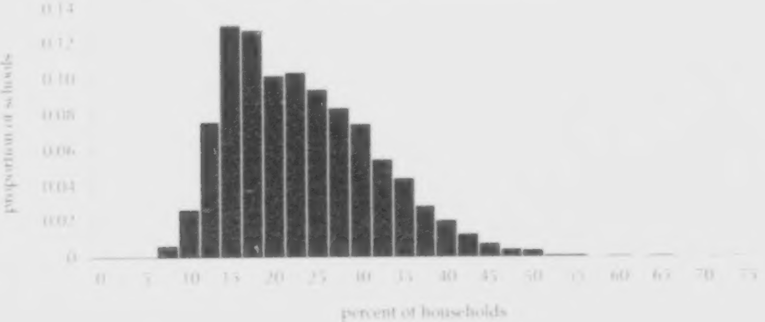
Note: The scatter plot uses data from schools with a total enrollment of 50 or more. In a very small school, the percentage of the school population in one category can be highly variable. There are 94 schools with an enrollment of fewer than 50 and 14 with an enrollment of fewer than 10. The plot is not substantially different when these schools are included.

school children live, on average about 45 percent of people over age 15 lack a high school diploma, while about 22 percent have at least some university education. Panel A of Figure 3.4 shows a fairly normal distribution of the percentage of adults without a high school diploma, but panel B shows a more skewed distribution of those with some university education. As Table 3.3 shows, only 318 of 4,055 Ontario elementary schools draw students from communities where more than 40 percent of people have some university education. A scatter plot of the two education variables — each dot in Figure 3.5 represents a school community — reveals that they do not have a simple linear relationship, which suggests there is useful information in both variables, a fact that becomes relevant in subsequent analysis.

Family Status Variables

Another useful census variable describing a school community is the percentage of single-parent families. As Table 3.3 summarizes,

Figure 3.6: ***Single-Parent Families in Ontario Elementary School Communities***



such families account for 22 percent of family units in the average elementary school community, although, as Figure 3.6 shows, the distribution of this variable is highly skewed, with a broad mass of school communities in which 15 to 30 percent of children live in single-parent families.

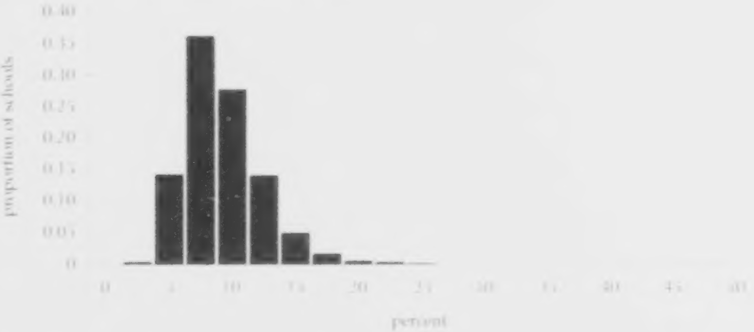
Economic Variables

The census provides the researcher with two variables of direct economic interest: unemployment rates and income measures. Figure 3.7 presents unemployment rates in Ontario elementary school communities for all people over age 15 and for people over 15 with children at home (again, descriptive statistics appear in Table 3.3). As the figure indicates, a few schools are in communities with very high unemployment rates, while average unemployment rates are lower when there are children present in a household.

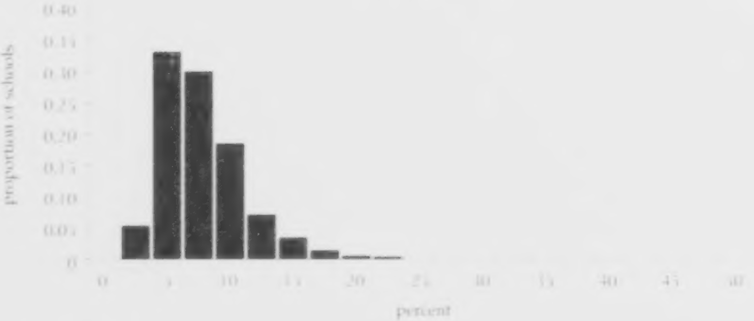
The census data contain four relevant measures of income: average family income for all families; average family income for two-parent families; average household income for all households; and average household income for households with two or more people. The four measures are, however, essentially identical in terms of the information they contain, as Tables 3.4 and 3.5 demonstrate. Accordingly, the measure I use throughout this study is average

Figure 3.7: ***Unemployment in Ontario Elementary School Communities***

Panel A: Unemployment rate for all adults



Panel B: Unemployment rate for adults with children



household income in households with two or more people, the school community distribution of which is presented in Figure 3.8.

Economists view income as a summary statistic. It obviously measures the monetary resources available to families, but it suggests other measures as well, such as fundraising opportunities at the school and the availability of additional educational resources outside the school system for children from higher-income families. Income also measures education, since the ability to earn a high income is strongly related to educational status, although the ability to earn a high income may also be a result of intelligence and good work habits. Thus, income may be a proxy for a variety

Table 3.4: *Measures of Income*

	Mean	Standard Deviation
	(1995 $)	
Average income, all census families	59,156	18,416
Average income, two-parent families	62,732	18,897
Average income, all households	56,785	18,092
Average income, households with two or more persons	62,017	18,773

Table 3.5: *Measures of Income, Correlation Matrix*

	Average Income			
	All Census Families	Two-Parent Families	All Households	Households with Two or More Persons
Average income				
All census families	1			
Two-parent families	0.99	1		
All households	0.97	0.96	1	
Households with two or more persons	0.99	0.98	0.98	1

Figure 3.8: *Average Household Income in Households with Two or More People, Ontario Elementary School Communities*

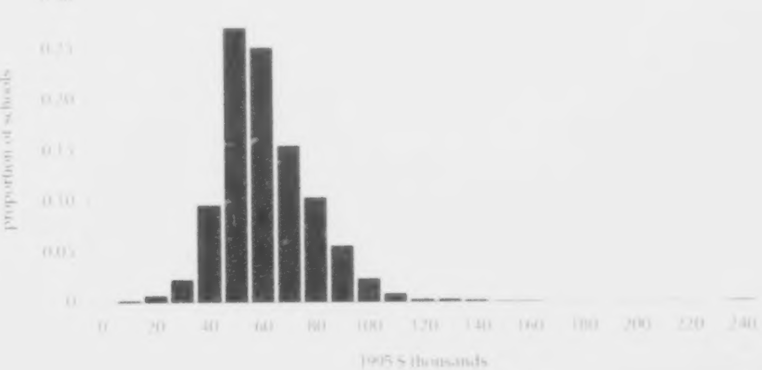
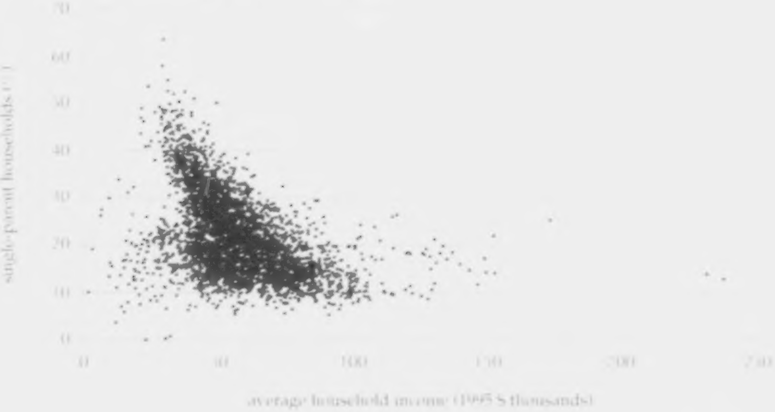


Figure 3.9: *The Relationship between Average Household Income and Single-Parent Families*



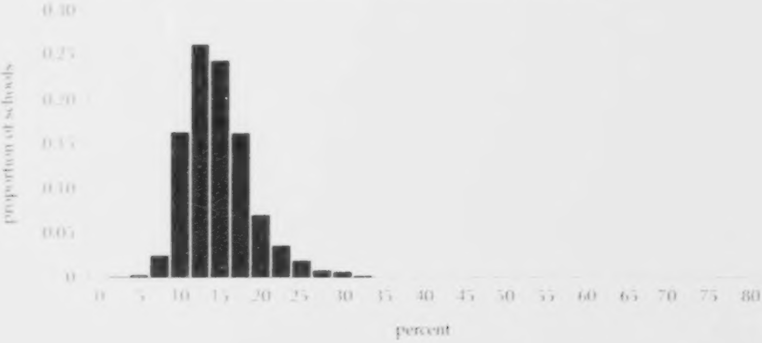
of unobserved characteristics of adults in the school community. In some cases, income may be inherited wealth. It is even possible for income to stand in for a measure of political influence — there is anecdotal evidence that, through political influence and the provision of more desirable teaching positions, schools in higher-income areas actually capture more or better resources per student than schools in lower-income areas. Moreover, while the household income measure that I use in this study is related to the structure of households, as is evident from Figure 3.9, there is enough difference between them to make both measures informative.

Mobility and Housing Variables

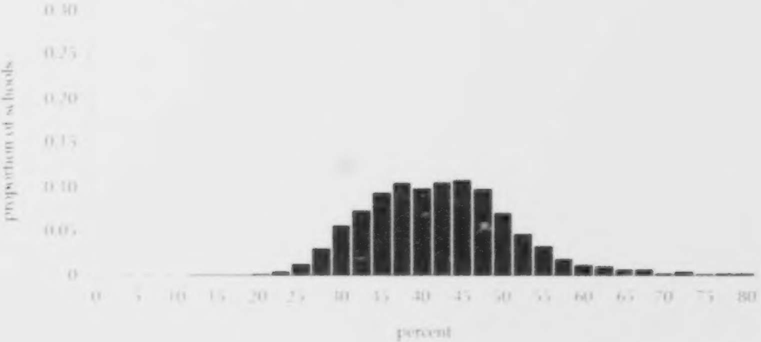
The census provides data on two mobility variables: the percentage of the population that moved within the past one year and the percentage that moved within the past five years (see Figure 3.10 and the descriptive statistics in Table 3.3). A very large proportion (more than 40 percent) of Ontario's population moves within any five-year period, so there is relatively little variation across school

Figure 3.10: ***Measures of Mobility in Ontario Elementary Schools***

Panel A: Percentage of the school community that moved in the past year



Panel B: Percentage of the school community that moved in the past five years

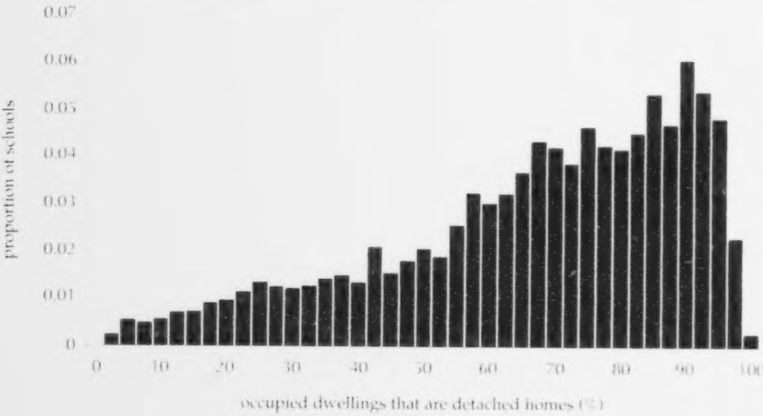


communities in this dimension. Moves within the past one year, however, are much more concentrated, with more than 20 percent of people moving in only 286 school communities.

The last useful variable from the census data is the percentage of single detached homes in the total number of occupied dwellings in the school community. The provincial average is 65 percent but, as Figure 3.11 shows, this number varies enormously across school communities.

Mobility and housing variables may also be considered economic in nature, but there is a very practical reason for presenting

Figure 3.11: ***Detached Homes in Ontario Elementary School Communities***



them separately. Unlike other census variables, they could be measured accurately each year without additional cost — that is, schools could easily identify their new students, those who had moved during the past academic year — and students taking the test could provide such information without the need for significant intrusion into their privacy.

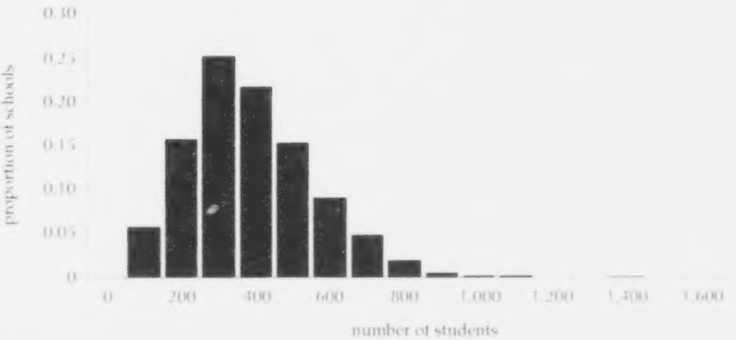
Variations in Enrollment and Sex

Ontario's postal code data allow two direct measures of the characteristics of the school itself rather than of the community from which it draws its students: total enrollment at the school and the ratio of boys to total enrollment (see Figure 3.12 for the distribution of these variables and Table 3.3 for the descriptive statistics).

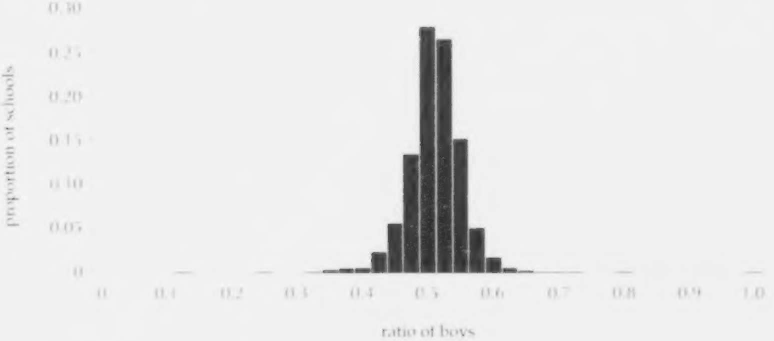
School enrollment is a useful variable because a systematic school-size effect on assessment results might suggest that there is an "optimum" size of school. Enrollment is also of interest because the test results of schools with fewer than 15 students in Grades 3 or 6 are suppressed for privacy reasons. Even where their test results are reported, smaller schools are likely to experience greater variations

Figure 3.12: ***Enrollment in Ontario Elementary Schools***

Panel A: Total enrollment



Panel B: Ratio of boys to total enrollment



not only in the results from year to year but also in their school community variables from one census to the next.

Of much less statistical significance in the assessment results is the variation among schools of the sex of students. This is perhaps surprising, considering the emphasis that some place on the underperformance of boys on tests.[3] Nonetheless, I include the variable to reassure the reader that its effect, however insignificant, is accounted for in the analysis that follows.

---

[3] The sex variation is subject to the error related to the complex duplicates I describe in the appendix to this chapter. The effect is not huge, however.

A Set of Interrelated Variables

The reader will no doubt have deduced that the variables I use to describe school communities are, to one degree or another, interrelated. Figures 3.5 and 3.9 have already presented the relationships between two pairs of variables; this section now takes a further look at the interrelationships among variables, summarized in the matrix presented in Table 3.6.

By way of explanation, a correlation coefficient is equal to 1 when the variable in the column has a perfect positive linear relationship to the variable in the row, and to –1 when the variable in the column has a perfect negative linear relationship to the variable in the row. The diagonal elements equal to 1 indicate that any variable has a perfect positive linear relationship to itself. A correlation coefficient of zero indicates that there is no relationship between the variable in the row and the variable in the column. The matrix is based on data for all of Ontario's elementary schools.[4]

Table 3.6 holds no real surprises. The two immigration variables are closely but not perfectly related, with a correlation coefficient of 0.89. The two language variables are closely related and have a strong negative relationship to the immigration variable. The two education variables, as suggested by Figure 3.5, have a strong negative relationship. A high value of the aboriginal variable is related to high unemployment, low income, and the presence of relatively few detached homes. The two unemployment variables are strongly related. There is a strong correlation between being unemployed with children and being a single parent. The household income measure is quite clearly related to both measures of education and has a weaker but still quite strong relationship to the measure of single-parenthood. Mobility is less when there is more detached housing but the relationship is not as strong as many other relationships in the correlation matrix. The transaction costs

---

[4] The correlation matrix is virtually identical when the 94 schools with an enrollment of fewer than 50 students are removed from the sample — that is, the relationships are not driven by outliers in very small schools.

Table 3.6: ***Correlation Coefficients among 16 Variables that Describe Ontario Elementary School Communities***

#	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	% of immigrants less than 5 years in Canada	1.00															
2	% born outside Canada, US, UK	0.89	1.00														
3	% with English or French as home language	-0.83	-0.87	1.00													
4	% with English or French mother tongue	-0.84	-0.95	0.95	1.00												
5	% with aboriginal status	-0.12	-0.26	0.11	0.20	1.00											
6	% without secondary school diploma	-0.11	-0.12	0.00	0.09	0.19	1.00										
7	% with some university ed	0.35	0.40	-0.24	-0.36	-0.30	-0.86	1.00									
8	% of children in single-parent families	0.37	0.29	-0.26	-0.24	0.08	0.24	-0.04	1.00								
9	% unemployment rate	0.37	0.25	-0.32	-0.23	0.34	0.42	-0.25	0.66	1.00							
10	% unemployment rate in families with children	0.45	0.32	-0.38	-0.31	0.26	0.39	-0.20	0.68	0.90	1.00						
11	Average household income	-0.07	0.07	0.11	0.03	-0.35	-0.70	0.71	-0.42	-0.58	-0.55	1.00					
12	% of households that moved in the past year	0.34	0.22	-0.17	-0.17	0.03	0.00	0.11	0.62	0.39	0.45	-0.21	1.00				
13	% of households that moved in the past 5 years	0.46	0.41	-0.27	-0.33	-0.16	-0.20	0.30	0.41	0.21	0.31	0.02	0.81	1.00			
14	% of dwellings that are detached homes	-0.67	-0.62	0.61	0.60	0.09	0.04	-0.26	-0.70	-0.47	-0.54	0.18	-0.60	-0.60	1.00		
15	Total enrollment	0.27	0.32	-0.21	-0.27	0.27	-0.20	0.25	0.02	-0.07	-0.03	0.23	0.10	0.30	-0.20	1.00	
16	Proportion of boys	-0.14	-0.20	0.14	0.19	0.10	0.11	-0.16	0.00	0.03	0.01	-0.15	-0.13	-0.24	0.10	-0.10	1.00

of moving from one unit of detached housing to another are much higher than those of moving between other forms of housing, so people who live in detached homes tend to move less often. There is less detached housing when there are more single-parent families, and more unemployment when average household income is lower. It is interesting that larger schools tend to have more immigrants and a larger proportion of families that move frequently. Finally, it is noteworthy that the correlation coefficients along the school size row and sex variation row are all relatively small.

The EQUIP Data

In the next two chapters, I place the test results and rates of exemption from tests in context using census data on a variety of socioeconomic measures that describe the school communities for more than 95 percent of Ontario's elementary schools. This analysis then leads to the value-added ranking of elementary schools that I present in Chapter 6. Before I turn to these tasks, however, I should let the reader know that EQAO, the Ontario government agency that conducts the tests, uses data from its Education Quality Indicators Program (EQUIP) to achieve similar goals. EQAO is fully aware of the need to place assessment results in the context of the characteristics of the local school community. To that end, it is engaged in the creation of a data set describing all Ontario public and separate schools, both elementary and secondary, although I am concerned here only with the variables created for elementary schools.

The EQUIP data have several advantages over the census variables I identified above. First, EQUIP includes a variety of administrative data not otherwise available to the researcher. Second, EQAO undertook a comprehensive consultation process as part of its choice of variables for the EQUIP data set. Third, EQAO is able to base some variables on the responses given by students who are actually being assessed, which provides data that describe the subjects of the tests — the students themselves — rather than the school as a whole.

The EQUIP data also have some disadvantages. One is their apparent lack of timeliness and openness: although the EQUIP program began some time ago, the raw data are still not in the public domain. Moreover, at least as far as the public description of the data on enrollments and postal codes are concerned, the EQUIP program uses data from the 1999–2000 academic year, which would omit a significant number of schools.[5] Another disadvantage, at least from the point of view of the researcher, is that EQAO itself has chosen the variables that are to relate to school performance; instead, we want to know which variables, from the available set of census variables, have the closest association with the assessment results.

EQAO has organized its chosen variables into themes: context, inputs, processes, and results (see Table 3.7). The context variables are similar to the variables I discussed earlier in the chapter. One difference, however, is the inclusion of an administrative measure of exceptional students — both those who are gifted and those who have other special needs — in the school; to the extent that the presence of exceptional students correlates to census variables that have already been already created, the census variables could proxy for such an effect. As for other context variables, the EQUIP mobility measure is the number of schools attended by students since Grade 1. This question, asked when the student participates in the Grade 3 or Grade 6 test, would need to be adjusted to be comparable across assessments, but it is a more precise measure of mobility than using the census and postal codes. The EQUIP income measure is constructed from postal code and census data. A parental education variable is also constructed from postal code data, although details of its construction are not given.

The input theme in the EQUIP data is addressed by, for example, asking parents about their children's pre-school attendance. There are also several measures of students' readiness to learn, measures of teacher and support staff inputs, student-to-teacher ratios, and teachers' qualifications. On the process theme, the EQUIP

---

[5] Updating may now be taking place, although the EQAO web site does not address this issue; see www.eqao.com/11e/11_1e.asp.

Table 3.7: ***The Elementary School EQUIP Variables***

Community, Student and Family Demographics

1. Student enrollment by grade
2. Family income levels
3. Parents' educational attainment
4. Student's language at home and length of residency in Canada
5. Students with special needs
6. Number of schools attended by student since Grade 1

Preparedness to Learn and Early-Learning Support

7. Attendance in nursery school prior to enrollment

School Leadership

8. School's indicators of leadership, planning, and decisionmaking

School Climate

9. Safety-related survey
10. Class sizes of Kindergarten, Grades 1-3, 4-8, and an indicator of split classes
11. Number of students per teacher and per non-teaching support person
12. Types of special education available

Community-School Relationships

13. Composition of school council and an index of activities

Teaching and Learning Environment

14. Time allocation to reading, writing, and mathematics
15. Accessibility of texts, other print, audio and video resources, and software
16. Availability of assessment materials
17. Availability of computers at home and at school
18. Frequency of teacher meetings to plan and discuss curriculum
19. Range of parental involvement
20. Measures of teachers' experience and qualifications

Student Achievement

21. Provincial achievement results in reading, writing, and mathematics

data include some survey results on school processes — such as school safety, attendance, and class size — which are solicited from schools and whose usefulness in relating to the assessment results is uncertain. Finally the EQUIP data will include the EQAO achievement results. If made public, the EQUIP data would be both a rich and useful dataset.

Chapter Summary

To place the results of standardized tests of elementary school students in context, one must determine the socio-economic characteristics of the community from which a school draws its students. Since privacy concerns do not permit a direct measure of each individual student's socio-economic characteristics, the next best method is to look at the average characteristics of households in the school community. For that, one must examine the data for households in small geographic areas as reported in the census. In addition, one can determine which of these geographic areas individual students live in by looking at their postal codes, which the Ontario Ministry of Education compiles for most of the province's elementary school students.

From the census and postal code data, one can glean an array of useful indicators of the school community's socio-economic characteristics, including the percentage of immigrants, both recent and long term, the language they speak at home, the number of aboriginals, the education of parents, the incidence of single-parent families, various measures of family and household income, the frequency with which families move, the percentage living in detached housing, the number of students enrolled in the school, and their sex.

Many of these variables are, of course, interrelated. For example, a high value of the aboriginal variable is related to high unemployment, low income, and the presence of relatively few detached homes. Unemployment rates tend increase with the number of single parents, household income increases with education, those who live in detached homes tend to move less often, there is less detached housing when there are more single-parent families, and more unemployment when average household income is lower. In addition, larger schools tend to have more immigrants and a larger proportion of families that move frequently.

Another set of socio-economic variables describing elementary school communities is compiled by Ontario's Education Quality and Accountability Office (EQAO) from its Education Quality Indicators Program (EQUIP). The EQUIP data have several advantages over the census variables: they include a variety of administrative data not otherwise available to the researcher, they were chosen, at least in part, following a comprehensive consultation process, and EQAO bases some variables on the responses of students who are actually being assessed. The disadvantages of the EQUIP data include their lack of timeliness and openness — the raw data are still not in the public domain, and the public data are several academic years out of date.

Technical Appendix: Constructing the School Communities' Socio-Economic Indicators

This appendix presents further details on the actual construction of school communities' socio-economic indicators, which should permit the interested reader to assess the quality of the data used in the study. The description should also be sufficiently comprehensive to allow other researchers to reproduce the results.

Step 1:

School Enrollment and Students' Postal Codes

There is no direct access to data describing the socio-economic characteristics of individual students. Fortunately, for the 1999–2000, 2000–01, and 2001–02 academic years, the Ontario Ministry of Education has kept files of the number and sex of elementary school students in each postal code in the province. This permits us to estimate the total enrollment at each school simply by adding up the students in all the postal codes that apply to each school community.

In undertaking this estimate, however, I discovered three small errors in the data.[6] First, the data for 1999–2000 and 2000–01 contain some exact duplicates — observations that are identical in every way — which I simply removed.

Second, the numbers of students in some postal codes were suspiciously large — for example, one postal code was listed as containing 2,000 girls but no boys, a very unlikely event. Upon checking with ministry staff, I discovered that these numbers were simple typographical errors, and I have now corrected them.

Third, the first two academic years of data contained some observations that I call "complex duplicates" — pairs of observations at the same school and postal code but with two different observations of enrollments. Such observations should not exist in the data. A typical example of a complex duplicate is where one observation lists an enrollment of, say, six boys and four girls, while the second observation lists an enrollment of six girls and four boys. Clearly, one of these observations is a data entry error. If such complex duplicates occur frequently at a specific school, it exaggerates the number of actual enrollments. To adjust for such errors, I removed one of the duplicate observations at random, since there is no way to know which observation is correct. This methodology should result in obtaining correct overall enrollment numbers for each school, although there may be some small errors in the actual numbers of students allocated by sex. Fortunately, while 4 percent of the data in the 1999–2000 file are exact duplicates and 0.2 percent are complex duplicates (with the errors highly concentrated at specific schools or boards), by the 2001–02 file such errors of duplication no longer appear.

---

[6] I discuss these errors not as a criticism of the data but as a warning to other users. I am grateful to Howard Kim at the Ontario Ministry of Education for his help in working with the postal code files.

Step 2:

Linking Postal Codes and the 1996 and 2001 Censuses

Responses to census questions — whether the small number that all Canadians are required to answer or the more detailed set that is asked of a 20 percent sample of the population[7] — are reported for a variety of geographic units, including very small ones known in the 1996 census as Enumeration Areas (EAs) and in the 2001 census as Dissemination Areas (DAs). EAs contain between 125 and 650 dwellings; DAs contain between 400 and 700 people. Although we do not know the characteristics of individual students who write the assessment at a school in a given year, knowing the student's postal code will tell us in what EA or DA the student lived. This, in turn, allows us to calculate various characteristics of an average person living in the geographic community that sends children to a particular school.

---

[7] For a description of the census data and forms, see Statistics Canada (1996, 2001).

There is a slight difference in how the 1996 and 2001 censuses link postal codes and the smallest geographic units. For the 1996 census, Statistics Canada provides a weighted postal code conversion file that signifies the proportion of the people in each postal code who live in a particular EA. For about 90 percent of the postal codes, the weight is 1, meaning that everyone in that postal code lives in a single EA. Some postal codes, however — particularly rural ones — cover a large geographic area that contains more than one EA. For such postal codes, Statistics Canada aggregates the EAs into an "artificial" EA using the weighted average characteristics of everyone living in the EAs linked to that postal code.

This exact exercise is not possible using the 2001 census. For that census, Statistics Canada provides a conversion file that links a postal code to one or more DAs. For 80 percent of the postal codes, one postal code is linked to one DA. In the other 20 percent of postal codes, there are links to several DAs. Unlike for the 1996 census, however, the 2001 postal code conversion file has no actual population weights. Instead, Statistics Canada provides a Single Link Indicator (SLI), which is set equal to unity if a substantial proportion of the population of a postal code lives in a particular DA. As a rough and ready guide, if only one such DA is tagged, all of the population in that postal code is assumed to live in that DA, and the other linked DAs are ignored. If two such DAs are tagged, a weight of 0.5 is placed on each DA for that postal code, and the other DAs are ignored. If three such DAs are tagged, each receives a weight of 0.333, and any other DAs linked to that code receive a weight of zero. And so on. Where there is an SLI for a code, all DAs for that code usually also have an SLI.

The weight of zero described above is unusual. Very occasionally, there is no SLI within a multi-DA postal code. The largest number of DAs associated with a postal code with no SLI is 83. In that and other similar cases, each DA receives a weight of 1 divided by the number of DAs. Using these various rules of thumb, a unique geographic unit is associated with each postal code in the 2001 census without the need to wait several years for the official 2001 postal code conversion file with the population weights. One encouraging outcome of this exercise is that the average weight on the conversion of a postal code to a DA in the 2001 census is virtually identical to the average weight on the conversion of a postal code to an EA in the 1996 census.

A third complication in linking postal codes to census units is that not all postal codes in the school data also exist in the postal code conversion file. In the most straightforward case — leaving aside typographical errors — the reason for the absence of some postal codes is that there are dwellings in the period covered by school postal code data (1999 to 2002) that had not yet been built at the time of the 1996 census. It is even possible that the 2001–02 academic year contains students who live in postal codes that are not part of the 2001 census, which was conducted in June of that year. The magnitude of this problem is not great, however: the 1990–2000 data contain six separate schools and the 2000–01 data just one separate school and one public school with students' postal codes that did not exist at the time of the 1996 census. If all the data from the 2001 census data were available, no schools would be excluded.[8]

We can also ask how many students are excluded from the sample because of a lack of census data on their school. A school could contain a mixture of new and old housing — that is, housing with and without census data. Some students at a school could be left out of the postal code data. Table 3.2 showed that, in the 2000–01 academic year, the combination of postal code data and the 1996 census reveals the average characteristics of the communities of 90 percent of public school students and 86 percent of separate school students. The 2001 census increases those values to 94 and 92 percent, respectively. By combining postal code and census data from the two censuses, the characteristics of the communities sending children to the vast majority of the province's elementary schools can be described.

---

[8] These comments help the reader interpret Tables 3.1 and 3.2.

Each postal code-EA (or postal code-DA) combination is weighted by the enrollment at that combination relative to the total enrollment at that school. The total school enrollment is determined by the sum of all the students living at census-linked postal codes. For example, suppose a school's students live in only two EAs, one with 60 percent of the school's enrollment and the other with 40 percent of the school's enrollment. If all the family units in the first EA are headed by single parents and but none of the families in the second EA is headed by a single parent, then the percentage of single parents in this hypothetical school community is 60 percent: (0.6 × 100 percent) + (0.4 × 0 percent).

Such measures can be constructed for the postal codes in each academic year and then for each census. A school could have from one to six measures representing the census characteristic of its community, depending on the availability of the postal code data at that school in a given academic year. To create the characteristic of the school's community, the different census-based measures from the different academic years and, where possible, the two censuses are averaged together.[9] It makes sense to average the census-based observations in any case, since the test results I examine in this study come from three academic years between the 1996 and 2001 censuses and from the academic year immediately following the June 2001 census. With one exception — the measure of family income — the units of measurement in the two censuses are the same.[10]

---

[9] A full set of measures based on only the 1996 census was constructed for an earlier version of this research, as described in Chapter 7, but the results are not sensitive to the choice of census.

[10] The 1996 census uses 1995 dollars and the 2001 census uses 2000 dollars. The second measure is deflated to 1995 dollars using the ratio of the Canada-wide consumer price index in 2000 to that in 1995.

4

Analyzing Test Results for Ontario's Elementary Schools

In its first provincial report, Ontario's Education Quality and Accountability Office (EQAO) stated that "[a]ssessment of student achievement is at the heart of our mandate" (EQAO 1997, 1). Analysis of student achievement is also at the heart of this book. In this chapter and in Chapter 6, I present two separate but closely related numerical analyses of EQAO's test results.

In this chapter, the unit of observation I use is the percentage of all students at a school achieving Level 3 or 4 (these are, essentially, the levels at which students meet or exceed curriculum expectations) on a specific reading, writing, or mathematics test in Grade 3 or 6 in a particular year. At most of Ontario's approximately 3,500 elementary schools with assessment data,[1] there is a total of 24 observations of achievement: one for each of the three subjects assessed, times two grades, times four academic years (that is, the four years since the tests were extended to cover Grade 6 in 1998–99).[2] There are, therefore, something like 84,000 observations in all.

---

[1] Counting the number of schools to be used in the analysis is actually quite complicated. Some schools have Grade 3 but not Grade 6, and vice versa. Some are French-immersion schools and thus were able to opt out of assessment in the non-immersion language in some years. In addition, schools opened and closed over the analysis period. In total, about 75 percent of the schools in the sample contain results from all years of assessments. The data set is thus quite untidy or, in the language of statistics, "unbalanced."

The chapter is organized as follows. In the next section, I describe the structure of the achievement data. In the following section, I present a comparison of year-over-year results on the same assessments and then a comparison of results within a year on reading, writing, and mathematics assessments. I then measure the relationships between the achievement data and the socio-economic factors that describe a school community, the factors I described in Chapter 3.

In next-to-last section, I explore the use of two small sets of factors that would be readily available at the level of the school or class at low cost. The first set uses the socio-economic factors EQAO has collected since the 2000–01 academic year for a sample of 10 percent of the schools in both 2000–01 and 2001–02. The second set restricts the census-based variables to those that a school could collect cheaply. Neither small set of variables, however, is as useful as the full set of data derived from the postal codes described in Chapter 3.

The chapter closes with a brief summary of what we have learned to this point in the analysis. A lengthy technical appendix presents the actual estimated equations and discusses some of the methodological issues around their estimation.

The Structure of the Assessment Data

Since the 1998–99 academic year, EQAO has conducted assessments of students in both Grades 3 and 6 in the areas of reading, writing, and mathematics. The test results actually contain more detail than is used in this analysis. Each assessment contains "strands" that identify areas of strength and weakness in a particular student or at a particular school. EQAO aggregates these strands into an overall score on reading, writing, and mathematics for each student.

---

[2] Chapter 6 presents an analysis of school-level achievement results averaged across assessments in Grades 3 and 6 in different years. This is a total of 7,000 observations, usually two observations at each school. Both types of analysis are useful.

EQAO then further aggregates these scores across all students at a school, importantly, not by class but *in a grade*. Most school contain both Grade 3 and Grade 6 classes; a few contain one grade or the other but not both. The school results present the percentage of students in each of the two grades, grouped in six or seven categories. (These percentages are rounded to the nearest whole number, so they may not add to 100 percentage points.) For the many schools in which there is more than one class in a grade, the class results are combined. If the school has fewer than 15 students in a grade, the results are suppressed for privacy reasons.

EQAO uses two methods to present the results of its assessments. Its "Method 2" relates to the exemption process and is described in detail in Chapter 5. Here, I consider EQAO's "Method 1," which places *all* students within a grade at a school in one of several categories. If sufficient material from a test is available to allow grading to take place, a student is assigned to Levels 1, 2, 3, or 4. As noted earlier, Levels 1 and 2 have come to mean that the student has not mastered the curriculum, Level 3 has come to mean that the student is meeting curriculum standards, while Level 4 now indicates that the student is exceeding curriculum standards.[3] In other categories are students who have taken the assessment but have not provided enough material to be graded, students (a tiny percentage) whose records have been lost in the system,[4] and, finally, the substantial group of students exempted from writing the assessments (which I consider in detail in the next chapter).

---

[3] I use this language because, in the first year of assessments, Level 2 was also considered to be satisfactory. However, the Ministry of Education now stresses that Level 3 and above is indicative of satisfactory performance.

[4] Students who do not provide sufficient material to be graded and those whose records are lost are not treated consistently across the four years of the data. In the first three years, the two are recorded as a single group; in the last year, they are recorded separately. This accounts for the variation in the number of categories of students.

In mathematical terms, the aggregated results can be displayed as follows:

Category One ($C^1_{S,i,t}$) = the percentage of students in the grade at school *S* in either Level 3 or Level 4 on assessment *i* in year *t*;

Category Two ($C^2_{S,i,t}$) = 100 (percentage points) minus the percentage of students formally exempt (EX) minus the percentage of students at Level 3 or Level 4 in that grade at school S in either Level 3 or Level 4 on assessment *i* in year *t* (this combines Level 1 and Level 2 students with students who provided insufficient material to grade and those whose papers were lost — the latter group is very small and not consistently recorded across the study years);

Exempt ($EX_{S,i,t}$) = the percentage of students in the grade at school *S* in either Level 3 or Level 4 on assessment *i* in year *t* formally exempt from all three tests.

The sum of the percentage of all students at Levels 3 and 4 — that is, Category One — has come to be regarded as the principal measure of the success of the elementary school system in Ontario, and is thus the measure I use in the analysis in the rest of this chapter.

A First Look at the Achievement Results

In view of the enormous number of observations of assessment results and school performance, it makes sense to look first for overall patterns in results, by year, and by type of assessment. Let us begin by examining Table 4.1, which shows the number of schools for which results are available each year and the percentage of students — averaged across schools — in each of the three categories noted above (Category One, Category Two, and the Exempt category),[5] and Figure 4.1, which shows the percentage of Grade 3 and Grade 6 students achieving Levels 3 and 4 in each of the three test subjects over the four-year period.

As the figure reveals, performance improved on the whole over the four years, but with some fluctuations.[6] For example, although the Grade 3 reading results (panel A) rose significantly from 1998–99 to 2001–02, the results were not significantly different in 2000–01 than they were in 1999–2000. Similarly, while the writing results for 2001–02 were the highest of the four years, they actually declined — slightly but statistically significantly — in 1999–2000 and 2000–01 compared with 1998–99. In mathematics, results increased over the first three years, then declined significantly in 2001–02 from the previous year.

---

[5] Some differences arise between the values in Table 4.1 and those in EQAO's Provincial Reports, for two reasons. First, the Provincial Reports present the percentage of all students involved in the assessments in each of the three categories, while my study excludes observations for schools in which there were fewer than 15 students in a grade. Second, a student at a small school is weighted more heavily than others in Table 4.1, while all students are equally weighted in the Provincial Reports. The differences are, however, small.

[6] As discussed in the technical appendix, looking at this data set involves an almost philosophical issue. In one sense, the data are *not* a sample — that is, they are conceptualized as including all schools and all students in the relevant grades. In this case, the values observed in Figure 4.1 are not estimates of the average results on assessments across schools, but the average results for all students in the grade across schools. In this sense, there is no sampling error and the concept of statistical significance is not relevant. In another sense, however, the data *are* a sample, since not quite all schools are included (although the sample is very large) nor are all students (those in other grades are excluded). I lean to interpreting the figures as related to taking a very large sample of students in these grades from a universe of students in all grades, but I recognize that interpretations may differ. If the reader prefers to treat the calculations as descriptive of the students in these grades and as a data set without sampling error, then all differences, no matter how small, are the changes in the measure of results.

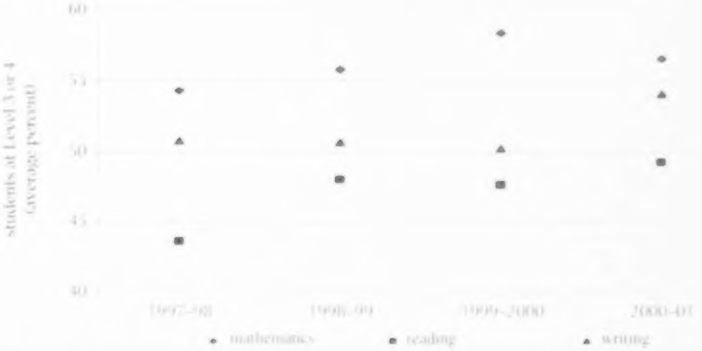
Table 4.1: ***Students in Each Assessment Data Category***

Academic Year	Number of Schools	Sum at Level 3 and 4	Sum at Level 1 and 2	Residual Category	Exemption Rate
		(percent, averaged across schools)			
1998–99					
Grade 3					
Reading	3,548	43.71	46.62	3.07	6.57
Writing	3,548	50.81	41.62	2.59	5.33
Mathematics	3,572	54.27	38.90	1.98	4.84
Grade 6					
Reading	3,339	48.57	44.32	2.59	4.37
Writing	3,343	47.66	44.02	4.19	4.15
Mathematics	3,339	46.41	47.26	1.99	4.19
1999–2000					
Grade 3					
Reading	3,567	48.00	39.71	5.38	6.90
Writing	3,567	50.68	38.57	4.75	6.00
Mathematics	3,592	55.82	36.37	2.44	5.37
Grade 6					
Reading	3,373	50.61	39.95	4.31	5.13
Writing	3,373	48.09	42.54	4.38	4.99
Mathematics	3,373	51.46	40.16	3.39	4.99
2000–01					
Grade 3					
Reading	3,498	47.67	38.45	8.01	5.97
Writing	3,498	50.20	37.44	7.10	5.32
Mathematics	3,521	58.36	31.57	5.46	4.68
Grade 6					
Reading	3,242	53.73	38.31	4.13	3.92
Writing	3,242	51.76	39.01	5.41	3.86
Mathematics	3,242	53.88	38.31	4.07	3.85
2001–02					
Grade 3					
Reading	3,341	49.29	36.25	7.38	5.62
Writing	3,334	54.06	36.16	3.54	5.08
Mathematics	3,370	56.57	31.87	5.57	4.44
Grade 6					
Reading	3,198	54.35	34.34	5.81	4.22
Writing	3,198	52.47	38.77	3.51	4.13
Mathematics	3,200	53.21	35.98	5.43	4.02

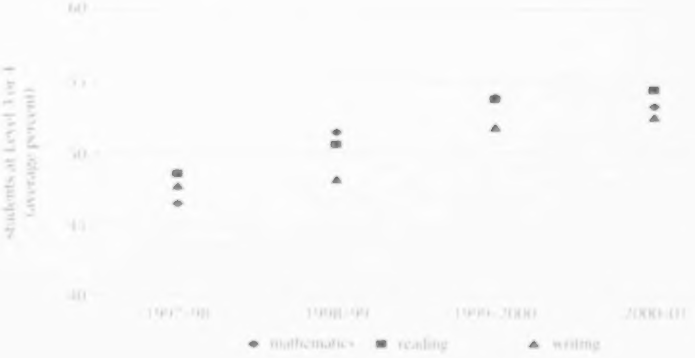
Source: Author's calculations using EQAO data obtained from a request under the *Freedom of Information Act*.

Figure 4.1: ***Trends in Average Achievement on Assessments, academic years 1997–98 to 2000–01***

Panel A: Grade 3 assessment



Panel B: Grade 6 assessment



The pattern of overall results on the Grade 6 assessment (panel B) is equally interesting, with statistically significant, and sometimes huge, increases in each of the three subjects from 1998–99 to 2001–02. The only exception is the increase in the writing test result from 1998–99 to 1999–2000, which was not significant.

It is also interesting to see how the average test results for each subject compare with those for the other subjects. In Grade 3, the mathematics results were always highest, followed by writing, then reading. In Grade 6, however, the order was not consistent: in the first year, the mathematics results were the lowest; then, in the middle two years, they were the highest (with reading right behind in 2000–01), but dropped to second place behind reading in the fourth year.

Figure 4.1 shows quite clearly that, on average, although test results for each subject vary considerably over the four-year period, the trend appears to be generally upward. The problem is that we really do not know if it is appropriate to interpret the results in this way since we do not know if each of the tests was of equal difficulty or was graded the same way. Readers will simply have to keep in mind that a higher value of Category One (students performing at Levels 3 and 4) in a given year could be the result of making the assessment easier or of giving a higher score to the same work. Since there is no analytical way to separate these two possibilities, that discussion is best left to the experts in the education field. In the most cheerful interpretation, however, Ontario's elementary students are getting better.

In contrast to this cheerful interpretation, then, what are we to make of the two disturbing exceptions to the steady increase in assessment results: the sharp drop in both the Grade 3 and Grade 6 mathematics results from 2000–01 to 2001–02? Again, it makes sense to regard higher average scores as an improvement in learning outcomes only if one is absolutely certain that the assessment instrument and its grading are consistent over time.

As a professor who has made up hundreds of examinations and tests over the years, I believe the likelihood of absolutely comparable assessments is quite low (as noted in Chapter 2, the Auditor General of Ontario has expressed a similar view). My reasoning is that, if the assessments were exactly comparable, then the application of common sense and the law of large numbers in statistics make it difficult to account for the two reductions in the overall mathematics scores. It is fairly straightforward to think of a situation in which an assessment was accidentally increased in difficulty or marked in a different way. It is hard, however, to imagine how large samples of similar children and teachers could suddenly reduce their performance from one academic year to the next. Did all the teachers involved suddenly become less skilled? Did they unlearn previously successful teaching techniques? Was the very large group of students assessed on knowledge accumulated over a number of years suddenly less competent on average than another very large group of students in the same grade in the previous year? For this reason alone, I conclude that using the assessment results to evaluate *relative* school success — the exercise I undertake in Chapter 6 — is the more promising use of the Ontario assessment process.

The reader should not, however, abandon all hope of knowing if the overall performance of Ontario's elementary school students is actually improving. It may be possible to address this question if the same assessment is given to students in different jurisdictions (other provinces, for example) at different times. If students in Ontario improved by more than those in other jurisdictions or if their scores declined by less than those in other jurisdictions on the same test (which would occur if the test were accidentally made harder), that would constitute evidence of improved performance on the part of Ontario's students.[7]

Neither should the reader assume from this discussion that the construction of Figure 4.1 and the associated analysis in the technical appendix to this chapter is a useless exercise. The figure, in fact, shows that variations both in average assessment results over time and across assessments in a given year are critical components of any analysis of the association of socio-economic factors with assessment results.

---

[7] This type of comparison is, in fact, possible as a result of assessments such as the School Achievement Indicators Project, the Third International Mathematics and Science Study, and the Programme for International Student Assessments, which are conducted across Canada in different provinces and in different years. There is a substantial literature comparing provincial results and differences in improvements across provinces on these assessments. They also have the advantage of being given in a number of jurisdictions at the same time. However, they cover only a small number of students within a jurisdiction and are not useful for comparisons across schools.

Identifying School Community Variables Associated with High Test Results

In Chapter 3, I identified a large number of socio-economic characteristics that describe the community from which a school draws its students, with a view to determining which of these characteristics are associated with higher assessment results at schools.

There is, however, a small complication to consider in the analysis of these types of data. In the "usual" situation in statistics, each observation of an event is considered to contain new information.[8] Suppose, for example, I want to estimate the average age of Ontario residents. I would begin by randomly sampling Ontario data from the census, which includes almost all Ontarians. As I increase the number of people in the sample, the age of each new person drawn into the sample would provide some useful new information about the exact average age of Ontarians. As the sample size increases, however, each additional piece of information on age becomes less and less useful since I can get a very exact estimate of the average age without using all the data in the census.

The key concept here is that each person I draw into my random sample provides equally useful additional information about the average age of an Ontario resident. However, this statement is *not true* as I observe additional assessments at the *same school* in trying to understand the relationships between the socio-economic characteristics of students and the schools' achievement results. As I noted earlier in the chapter, there are — at least, there appear to be — approximately 84,000 observations of the effect of school community characteristics on school achievement. But even though each observation contributes a little bit to our knowledge of these relationships, this number is misleading. A specific school shares, over all its assessment results, the context information that describes the community in which it is situated,[9] but not all of the information from each observation is "new." Thus, we do not have 84,000 *independent* observations.

---

[8] I discuss this statistical issue at length in the technical appendix to this chapter, but also outline it here so that readers who wish to skip the appendix may still gain an intuitive understanding of the nature of the data on school achievement.

The technical appendix to this chapter takes into account this organization of the data in estimating the relationships between school community characteristics and assessment results. Here, I describe these relationships only in general terms. However, even in a "nontechnical" presentation, there are, unavoidably, two equations to discuss. The first is:

$$C^1_{S,i,t} = \alpha + \delta_{i,t} D_{S,i,t} + \beta_X X_S + u_S + u_{S,i,t}. \qquad (4.1)$$

This is a much simplified version of the actual equations estimated in the technical appendix. It describes the value of Category One, the percentage of students performing at Level 3 or Level 4 at school S on assessment i (reading, writing, or mathematics in either Grade 3 or Grade 6) in academic year t (1998–99, 1999–2000, 2000–01, or 2001–02). The coefficient α is a constant. The coefficients denoted $\delta_{i,t}$ are each multiplied by a "dummy" variable, $D_{S,i,t}$, that takes on a value of 1 when the assessment at school S is of a specific type in a specific year, and zero for all other assessments.

If we look only at the Grade 3 tests, there are 11 such dummy variables. If we look at the Grade 6 tests, there are also 11 such dummy variables. Notice that there is always one less dummy variable than the total number of assessments in Figure 4.1. The reason is that each coefficient on each of the 11 dummy variables in each equation has a clear interpretation: the coefficients measure the change in the percentage of students achieving at Level 3 or Level 4 on the assessment which that dummy variable represents relative to the assessment that does not have a dummy variable in each equation. The assessments are grouped by both type (reading, writing, mathematics) and year (1998–99, 1999–2000, 2000–01, 2001–02). Figure 4-1 reveals the importance of allowing the different assessments in different years to have different average results. If the value of $\delta_{i,t}$ is 3.02 on the dummy variable that represents the Grade 3 mathematics assessment in 2000–01 when the Grade 3 reading assessment in 1998–99 is represented by the omitted variable, then the average achievement level on Grade 3 mathematics in 2000–01 is 3.02 percentage points higher than the average achievement level on reading in 1998–99. These are the effects we see in Figure 4.1, and are included "in passing" in estimating equation (4.1). They are not, however, the effects of central interest in our analysis.

---

[9] The EQAO website (www.eqao.com) contains some class-specific data on the context of a year at a school. A sample of that data is used below as a check on the census data.

The main problem is, rather, to measure the relationship between the school community's socio-economic characteristics and the school's achievement results. Imagine that, in equation (4.1), X_S is one such variable at school S.[10] (Notice that the value of X at school S does not vary by year or type of assessment.) We simply know the actual value of the school community characteristic at that school. The value of the coefficient β_X in equation (4.1) is a measure of the association between the school achievement result, the value of Category One in an assessment at a school, and the value of the variable X at all schools.

To be concrete, suppose that X_S represented average household income in dollars, the value of which varies a great deal from one school to another. Remember that $C^1_{S,i,t}$ is the percentage of students at Level 3 or Level 4 on assessment *i* in year *t* at school *S*. Thus, if the value of β_X was 0.0002, then, as average household income at a school increased by $10,000, the average value of C^1 for all schools in that year for that assessment would be expected to rise by 2 units — the units being percentage points of students at Level 3 or Level 4. In other words, $10,000 in additional household income in a school community would be associated with, on average, a 2 percentage point increase in the students achieving at either Level 3 or Level 4. If school number 100 (that is, *S* = 100) had an average household income $10,000 higher than school number 122 (that is, *S* = 122), one could predict that, on average, achievement scores at school number 100 would be 2 percentage points higher than those at school number 122.

---

[10] X_S could represent many variables, as I show in the actual estimation of equation (4.1) in the technical appendix. I explore the two-variable case shortly.

The characteristic X_S need not be household income. It could be the percentage of children living in a single-parent household, the percentage of the school community living in detached homes, the size of the school, or any of the other measurable characteristics of a school community. The technical appendix explains how the value of β_X is estimated using an ordinary least-squares regression of equation (4.1).

Two vital results can be obtained from this analysis. First, it can show which school community descriptive variables are actually associated with higher or lower assessment results. Second, the value of the sum of u_S and $u_{S,i,t}$ can be determined for a typical assessment at a typical school.

Looking back at equation (4.1), the sum of u_S and $u_{S,i,t}$ is the part of the variation in a specific assessment result at one school relative to another school that is *not* explained by the variation in the school's socio-economic characteristics *nor* by the variation in the average result across the province on that assessment in that year. To determine how much of the variation in achievement results is associated with variations in the socio-economic variables that describe a school community, one must look at the R-squared (R^2) when equations such as (4.1) are estimated. A visual picture of R^2 is found in the technical appendix in Figures A4.1 and A4.2. If an R^2 is unity or close to 1, then *all* or *nearly all* of the variation in assessment results is associated with variations in the school's socio-economic context.

One might suppose that the average absolute value of the sum of u_S and $u_{S,i,t}$ is very small. In the estimates of equations such as (4.1) and, as we shall see, (4.2), however, this is simply not the case. In the technical appendix, the highest values of R^2 indicate that only about 25 percent of the variation in achievement results across schools, grades, assessments, and years is associated with variations in school community socio-economic variables. Thus, the average absolute value of the sum of u_S and $u_{S,i,t}$ is quite large, and fully

75 percent of the variation in assessment results in a given year on a given assessment is not explained.

In one way, this result is not surprising. Kane and Staiger (2001) show, in similar equations with individual student assessment results, that elementary school classes essentially are collections of individuals who vary enormously from year to year, and that socio-economic descriptions of them cannot fully explain variations in their achievement results. Such factors are important in explaining strong assessment results, but they are not the only explanation.

In another way, however, it is surprising that the R^2 is so small when the equations are estimated using class groupings at the same school. Many people would say good achievement results come *only* from schools with high-income and well-educated families and that the variation in achievement results is mostly due to those factors. At the same school, socio-economic factors do not vary much from one year to next and the socio-economic characteristics of many students are similar. Thus, if assessment results were very strongly related to school context, one might expect to see an R^2 higher than 0.25. Thus, it is demonstrably wrong to suggest, as some do, that school assessment results reflect *only* the socio-economic characteristics of a school's students.

I revisit this important issue in Chapter 6 in another context. For the rest of this chapter, however, I focus on determining which of the socio-economic variables that describe the school community are most closely associated with high or low assessment results.

Equation (4.1) contains just one context variable, X_S, that describes a characteristic of the school community. In reality, there are many context variables, a long list of which I presented in Chapter 3. For example, if there were two context variables, then the estimated equation would be:

$$C^1_{S,i,t} = \alpha + \delta_{i,t} D_{S,i,t} + \beta_X{}^1 X^1_S + \beta_X{}^2 X^2_S + u_S + u_{S,i,t}. \qquad (4.2)$$

The interpretation of this equation is similar to that of equation (4.1). In the previous example, X^1_S was the average level of household income in the school community. Now, suppose that X^2_S is the percentage of single-parent families in the school community. If

$\beta_X^2 = -0.10$, then a 1 percentage point increase in the percentage of single parents *at the same level of household income* would be predicted to reduce average assessment results, as measured by the value of Category One, by 0.1 of a percentage point.

The analysis need not be limited to two context variables; indeed, in the technical appendix, more variables are added as required. The only issue is whether the additional variable that describes the school community has a strong statistical association with higher or lower assessment results after taking into account the association of other variables.

One way of thinking about the addition of additional variables is to ask whether they reduce the average absolute value of u_S and $u_{S,t,p}$, the variations in results that are *not* associated with variations in socio-economic variables. Although detailed tables are presented in the technical appendix, Table 4.2 summarizes estimates of equations from both the Grade 3 and Grade 6 assessments to answer the question: What socio-economic variables, when measured at the school level, have a strong association with higher or lower results on assessments? Using the estimated coefficient β_X, the table shows the size of the predicted average effect on school achievement levels of a one standard deviation change in the value of that school community characteristic in both the Grade 3 and Grade 6 assessment. Changing a variable by one standard deviation is a way of making comparable changes across variables, otherwise one would have to compare unrelated variables such as increases in the percentage of single parents and increases in the percentage of students speaking an official language at home.

The data in Table 4.2 present several clear results.[11] Looking first at the Grade 3 achievement results (the percentage of students at either Level 3 or Level 4), we see that, in school communities with a higher percentage of single-parent families, results are usually lower, *all other factors being the same*. To see how this result is derived,

---

[11] As I explore further in the technical appendix, the estimates in Table 4.2 assume that the effect of a socio-economic context variable is the same on all three assessments. This assumption is usually valid.

Table 4.2: *The Association of School Community Socio-Economic Variables and Grade 3 and Grade 6 Assessment Results*

School Context Variable[a]	Mean[b]	Standard Deviation	Grade 3 Effect[c]	Grade 6 Effect[d]
Single-parent families (%)	22.0	8.6	-0.95	-1.26
Official language spoken at home (%)	91.5	9.5	-0.95	-0.47
Aboriginal descent (%)	1.6	2.6	-1.49	-1.25
Recent immigrants (%)	4.3	5.6	0.71	0.67
Born outside Canada, U.S. or U.K. (%)	19.1	17.8	0.68	0.02
Living in detached home (%)	65.7	23.1	1.39	0.76
Logarithm of household income[d]	10.9	0.3	1.97	2.11
Moved in past year (%)	13.1	4.0	-1.56	-0.80
Moved in past 5 years (%)	41.1	9.4	-0.66	-0.65
Unemployment rate (%)	6.5	3.4	-0.90	-1.09
Adults without high school (%)	45.1	9.7	-2.14	-0.86
Adults with some university (%)	21.9	11.1	2.55	5.66
Total enrollment	350	165	0.5	-0.17
Total enrollment squared	150,329	151,434	-0.90	1.35
Number of assessments[e]			41,122	38,762
Number of schools			3,403	4,111
R^2 in equation used			0.26	0.28

[a] The variable descriptions are found in Chapter 3.

[b] Mean and standard deviation are from Table A4.3. Samples for Grade3 and Grade 6 are not identical.

[c] The effect is calculated from regressions described in the technical appendix to Chapter 4. The value is the average effect, across schools and assessments, of a one standard deviation change in the associated variable on the Grade 3 or Grade 6 assessment result.

[d] Using the logarithm of income allows a nonlinear effect in the association of income and achievement results. This is important. Intuitively, increasing average household income at a school from $25,000 to $50,000 is more important than increasing average household income from $50,000 to $75,000. Using the logarithm of income allows this to be the case. More detail is provided in the technical appendix to Chapter 4. The change considered in the income variable is from the average household income in the sample of $54,176 to $71,682 (both in 1995 dollars). This change is predicted to increase assessment results by 1.39 percentage points in Grade 3 and by 2.11 percentage points in Grade 6.

[e] Each school may have up to 12 assessments in a grade. Not all schools have data available for all years and not all schools exist in all years. Not all schools have both Grade 3 and Grade 6.

note that 22 percent of the families in the average elementary school community are single-parent families. The standard deviation of the single-parent school community variable is 8.6 percentage points.[12] So, at a school whose community contains 8.6 percentage points more single parents than another school, the Grade 3 achievement score should, *all other school context variables being the same*, fall by 0.95 of a percentage point. This is a substantial effect.

Table 4.2 also shows that, perhaps surprisingly, if a larger percentage of the school community consists of immigrants who have come to Canada within the past five years, achievement scores tend to be higher. Most people might think recent immigrants would find standardized tests difficult, but immigrants are far from random samples of individuals from their own countries. Rather, they tend to be motivated people who are willing to move to improve their situation and provide better opportunities for their children. And their children do well in school when other factors are taken into account.[13] Indeed, they apparently do so well that, if a larger percentage of the school community speaks one of the official languages at home, then achievement results tend to be lower. When the immigrant characteristic of the school community is measured as the percentage of all residents born outside North America and the United Kingdom, any association with higher or lower assessment results is weaker. Of course, some of these immigrants may have been in Canada for a long time and their children may have been brought up in this country.

Table 4.2 also shows, perhaps not surprisingly, that the Grade 3 achievement results tend to be lower when a larger percentage of the school community is aboriginal.

---

[12] This means that about 66 percent of school communities have between 22 – 8.6 = 13.4 and 22 + 8.6 = 30.6 percent of single-parent families.

[13] A cheerful interpretation of these coefficients is that the Ontario school system meets the needs of immigrant children very well. At a minimum, the coefficients suggest that immigration and language characteristics of schools are not dominant context variables in their association with higher or lower school achievement results. The association of language and immigrant status is weaker with the Grade 6 results than with the Grade 3 results.

The three economic variables that describe the school community all have the expected strong effects: achievement levels rise as detached housing and average income increase, and fall as the unemployment rate rises.

The income variable raises a number of additional issues. As with all the other variables, income is used simply to describe its association with higher or lower achievement results without implying that it is the *cause* of higher or lower achievement results. Indeed, it is always difficult to state what the coefficients on the variables in these estimated equations actually mean in terms of causation. The interpretation of the coefficient on income, however, creates particular problems. True, higher household incomes are strongly associated with higher achievement results,[14] but the economist views the income variable not as predetermined but as a result of choices and skills. Clearly, higher incomes are partly the result of higher educational levels of parents in the school community. But education levels appear independently in the equation, so the coefficient measures the effect of higher income after education levels have been taken into account. Higher incomes may also reflect more innate ability on the part of the "average" adult in the school community, some of which would likely be passed on to children. Higher income, all else being equal, is also likely associated with the provision of more learning resources for children outside the school itself. There may be more resources in the home, more extracurricular activities, more effective school fundraising, or more pre-school preparation. A higher average income may create peer effects in the school, whereby students from homes with more economic advantages share resources with students from homes with fewer economic advantages. From the point of view of comparing school assessment results and similar schools (the Chapter 6 project), the reasons for the positive relationship between higher income and higher achievement results are not central. From the point of view of understanding why higher income is associated with higher assessment results, the reasons for this association are central. Clearly, more study is needed on the interpretation of this coefficient. It is enough for the main purposes of this book, however, that the coefficient is large and highly statistically significant, but that the inclusion of income in equations such as (4.1) and (4.2) does not lead to an R^2 of 1.00.[15]

---

[14] As discussed in the technical appendix, the income variable is actually the logarithm of household income in order to allow for a nonlinear relationship between assessment results and income.

Table 4.2 also shows, as expected, a very strong association between test results and the variables that measure the educational background of adults in the school community, even after taking into account variations in income. Notice that there are two education variables, which allows us to examine the strong effects of the extremes of the education spectrum on school achievement results. Thus, a 9.6 percentage point increase in adults without a secondary diploma results in a 2.13 percentage point fall in test scores, while an 11.1 percentage point increase in adults with some university education results in a 2.55 percentage point rise in test scores. A school community that reduces the percentage of its adults without a high school diploma by 10 percentage points and increases those with some university education by a similar amount (a plausible change in a school community in Ontario) would see achievement results rise by about 5 percentage points. This is a very large change in the expected level of achievement.

Both of the variables that measure the mobility of the school community have an important statistical association with the level of achievement on the Grade 3 assessments. An increase in either the percentage of the school community that moved in the past year or in the past five years reduces achievement results. The first effect is much stronger than the second.

---

[15] Remember that, if the R^2 were 1.0, then all variations in test results would be explained by variations in household income. To take an extreme and silly example, suppose that income were perfectly related to ability, that ability were entirely passed on through genetics, and that test results were a perfect measure of ability. Able parents would then have able children, and no matter what happened at the local school, all variations in test results would be related to parental income as the numerical counterpart of family genetic ability. This is obviously not the case. The unequivocal message of the interpretation of these test scores is that what happens at the local school matters!

The final variables included in Table 4.2 and the associated tables in the technical appendix are two measures of school enrollment — that is, of school size: the level of total enrollment and the level of total enrollment squared (which allows school size to have a nonlinear effect). These variables are included because of the controversy in education about the effect of school size and the closure of small schools. In fact, however, the effect of the level of total enrollment on achievement is not statistically significant. There is a negative and statistically significant coefficient on the square of enrollment, but the total effect is not of much practical interest. The average school in the sample has 350 students, and the standard deviation of school size is 165. This would, in the model, increase achievement results in the level of the school size by 0.50 of a percentage point and reduce achievement results in the square of school size by an approximately offsetting 0.85 of a percentage point. Thus, school size is not a decisive variable in school assessment results.

Having gone through the Grade 3 results in Table 4.2 in some detail, we can go more quickly through the Grade 6 results. There are some important differences and many similarities between the results for the two grades. The association between single parenthood and lower assessment results is similar in magnitude in both grades, as is the association between aboriginals and lower scores. In the Grade 6 version of equation (4.2), however, the coefficients on the variables that measure the percentage of the school community that speaks an official language at home, on recent immigrants, and of all immigrants are no longer of statistical significance and they are smaller in absolute magnitude. That is a very interesting result. Ontario's school boards, in their handling of students from a language background other than that of an official language, create a situation where, by Grade 6, those students have a success rate in literacy-based assessments equal to that of students who speak an official language at home.[16]

---

[16] As the technical appendix shows, schools with larger populations drawn from recent immigrants actually substantially outperform other schools on the Grade 6 mathematics assessment.

The role of economic variables continues to be strong in the Grade 6 results. The association of housing status measures and household income measures with Grade 6 assessment results are quite similar to those for Grade 3. Higher average incomes in the school community are strongly associated with higher scores, while higher unemployment rates are associated with lower scores. If there are more detached homes in the school community, then assessment results in Grade 6, as with Grade 3, tend to be higher. Increases in the one-year and five-year mobility variables are associated with lower scores, as they are for Grade 3, although the relationships are slightly weaker.

An interesting and important change between Grade 3 and Grade 6 is in the association of parental education and assessment results. By Grade 6, an increase in the percentage of the school population without a high school diploma no longer seems to have a significant impact on test scores: increasing by one standard deviation the percentage of adults without a high school diploma reduces the Category One variable by 2.13 percentage points for Grade 3 but only by 0.86 of a percentage point for Grade 6. At the same time, however, the percentage of the adult population in the school community with some university education has a much stronger association with higher test results in Grade 6 than in Grade 3: increasing by one standard deviation the percentage of adults in the school community with some university raises assessment results by 2.55 percentage points for Grade 3 but by 5.66 percentage points for Grade 6.

As with the Grade 3 results, school size measures have no substantial effect on the Grade 6 results. If anything, larger schools tend to have slightly better results on the Grade 6 results.[17]

---

[17] The alert reader may have noticed the absence of sex variables in the equations. The available sex variable — the proportions of boys and girls at each school averaged over the three years of postal code data — is *not* the same as the proportions of boys and girls in the grade at the school writing the test, and was never statistically significant in equations (4.1) or (4.2). It has accordingly been excluded from the analysis.

Before moving on, there is one other important question to discuss. Do the socio-economic characteristics of the school community have the same association with higher or lower scores regardless of whether the assessment is on reading, writing, or mathematics? In estimating equations (4.1) and (4.2), I have made this assumption, but is it correct? For example, it may be the case that students from higher-income families do much better in mathematics than those from lower-income families, but not in reading. Perhaps the association of single-parent families with test results on reading, writing, and mathematics is unequal. Perhaps parental help is more crucial in mathematics than in reading and writing. In fact, these and similar assumptions are tested in the technical appendix, which yields a few interesting results.

Speaking an official language at home is strongly associated with lower Grade 3 scores in mathematics. Having many adults in the school community with some university education also lowers Grade 3 mathematics scores even beyond the effect on reading scores. Aboriginals show a particular tendency to have lower scores in mathematics in both Grades 3 and 6, while recent immigrants do well in mathematics. In contrast, there is relatively little differentiation in the association of socio-economic variables with higher or lower scores in reading versus writing, which is what one would expect, since presumably reading and writing skills are more closely related to each other than they are to the skills needed to master mathematics.

Are the extra interaction variables worth adding to the explanatory equation? In short, no. The increase in the equation's explanatory power with the addition of these variables is tiny. As the technical appendix shows, the increase in the R^2 occurs only in the third decimal place. Accordingly, although there is further interesting work to be done in the analysis of these interaction effects, I omit these effects from the estimation process in the rest of the book.

"Cheap and Easy" Sets of Context Variables

Is it possible to use much less complicated sets of context variables than the ones I have so far described to predict assessment results?

For its part, EQAO is fully aware of the need to provide context for the achievement results of its tests; accordingly, in reporting results for the 2000–01 and 2001–02 academic years, it includes just four variables that describe characteristics of the students who have written the tests:

- the percentage of students enrolled in English-as-a-second language (ESL) programs;
- the percentage of students receiving special education support;
- the percentage of students that speaks only or mostly a language other than English at home; and
- the percentage of students that speaks a language other than English as often as English at home.

From the data EQAO has posted on its web site, I created a random sample of 350 schools — every tenth school from an alphabetical list. I then estimated equation (4.2) for the sample and the EQAO context variables (the estimated coefficients for these relationships are presented in Table A4.6 in the technical appendix). The most interesting result of this analysis is that, even with the advantage of data that directly describe the students writing the assessment rather than data related to characteristics of the whole school community, the EQAO context variables do not "explain" as much of the variation in assessment results as the variables created from the postal codes and census.

Nevertheless, the EQAO variables display some interesting interactions. For example, when more students speak the language of instruction (either English or French) at home, the average assessment result falls. The effect is statistically significant in the Grade 6 equation but not in the Grade 3 equation. This is consistent with the counterintuitive results I reported above, where a larger proportion of immigrants is associated with higher scores after the language effect has been taken into account.

Another interesting result is that, as the percentage of students in an ESL program increases, the assessment results tend to fall, with the effect much more pronounced on the Grade 3 test than on the Grade 6 test. Then, after controlling for the direct effect of enrollment in an ESL program, assessment results tend to rise as languages other than the language of instruction are spoken at home.

A much less interesting result is that, if a class contains more special needs students, this is associated with lower test scores. This effect is very large, but since the Category One measure is the percentage of *all* students in Level 3 and Level 4, and given that many or most special needs students are exempted from taking the tests, then the percentage of all students that achieves Level 3 or Level 4 as a proportion of the class must fall as a matter of arithmetic.

In the technical appendix, equation (4.2) is also estimated with a second small set of five context variables from the census and school. This time, only the one-year mobility measure, a language variable, the percentage of the school community that lives in detached housing, school size, and school size squared were included in the equation. The logic is that these five variables could be collected directly from students or the school at low cost without significant intrusion on privacy. Using this small set of variables in equation (4.2), as Table A4.7 in the technical appendix shows, does not successfully explain the variation in school assessment results.

In summary, neither the four EQAO variables nor a set of "cheap and easy" variables from the census provide a satisfactory set of context variables. It is, in fact, necessary to have a multidimensional view of the socio-economic context of the school community in order to predict school assessment results.

Conclusions

At this point in the analysis, one can venture four conclusions.

First, average results on assessments vary by year and type of assessment.

Second, the socio-economic variables that describe the characteristics of the school community, as well as the dummy variables that control for differences in mean results by assessment, do not explain all the variations in achievement results — there are important residual variations across schools within a year, across years at the same school, across assessments within the same year, and across years. Most important, different schools take students with similar backgrounds and produce different results, a phenomenon I explore further in Chapter 6.

Third, although the relationships between socio-economic variables that describe the school community and the assessment results on the Grade 3 tests and those on the Grade 6 tests are very similar, they are not identical. It is thus important to analyze the Grade 3 and Grade 6 results separately.

Fourth, "cheap and easy" data sets that attempt to place school assessment results in context do not work as well as the larger and more complex data set based on census and postal code data.

In the end, placing the school assessment results in context requires an effort. It is a substantial, but not impossible, task. In the next chapter, I analyze the relationships between the exemption rate and the variables that describe the school community. Then, in Chapter 6, I use a variant of the estimated relationships between the achievement data and the socio-economic variables to identify specific schools that outperform schools with equivalent socio-economic characteristics.

Chapter Summary

This chapter and Chapter 6 are the heart of the numerical analysis of school-level EQAO assessment results in this book. In this chapter, each of the (up to) 24 assessment results, the percentage of all students at the school in either Grade 3 or Grade 6 achieving Level 3 or Level 4 on a reading, writing, and mathematics assessment in any of four years, is an observation of achievement at a school.

The analysis of these data reveals a number of important features. First, average achievement results increased across the province over the four-year period. In both mathematics and the two measures of literacy, the average value at a school of this measure of achievement was substantially higher in the 2001–02 academic year than in 1998–99.

It is difficult to interpret this fact. One major concern is test scores did not increase between *each and every year*, a pattern that would have been expected if the increase in scores reflected long-term improvement in the educational system. A higher achievement result could also have occurred if an assessment had been made easier or graded differently. Since average assessment results in mathematics fell in both Grade 3 and Grade 6 from 2000–01 to 2001–02, one has to decide how to interpret that result. Common sense suggests that it is much more likely that the assessment varied in difficulty or in grading than that an immense number of students and teachers suddenly changed their characteristics for the worse in both Grade 3 and Grade 6 from 2000–01 to 2001–02. Results also varied across assessments within a year. The Grade 3 results were usually higher in writing and mathematics than in reading, while the Grade 6 writing results were lower than the reading results in all years. In some years but not others, the Grade 6 mathematics results were higher than the reading results. Part of the analysis of school results must take into account the variation in average results over time and across assessments.

The second part of the chapter measures relationships between assessment results and a large variety of context variables that describe the community from which a school draws its students, as well as school enrollment (size). In the analysis, I pose five questions and obtain the following answers.

First, *what percentage of the variation in assessment results is associated with variations in the characteristics that describe a school's community?* The answer is that, depending on the exact set of variables chosen, about 25 percent of the variation in school assessment results is associated with either measurable characteristics of the school community or variations in the average assessment result by year or type. That leaves a substantial amount — 75 percent — of unexplained variation that is not associated with observable characteristics of school communities or variations in the average test result. Two school communities with identical measurable characteristics may have very different results in a given year or over a number of years. The implications of this result receive further attention in Chapter 6.

Second, *what variables that describe a school community are most closely associated with higher or lower results on assessments?* There are only a few surprises. School size, for example, appears to have no substantial effect. Higher achievement results are strongly associated with school communities with higher incomes and higher education levels, but other socio-economic characteristics are also associated with higher or lower achievement results. If there is a large percentage of single-parent families, of aboriginals, of households that move frequently, of people that are unemployed, or of people that live in nondetached housing in a school community, then achievement results tend to be lower.

Third, *are the same school community variables associated with higher or lower assessment results on each type of assessment — that is, reading, writing, and mathematics?* The answer here is yes — there is very little variation in the socio-economic variables associated with higher or lower achievement results by subject.

Fourth, *are the same school community variables associated with higher or lower assessment results in both Grade 3 and Grade 6?* Here, the answer is no — the socio-economic factors associated with stronger Grade 6 results are, in fact, slightly different than those associated with stronger Grade 3 results.

Fifth, *is it possible to use small, "cheap and easy" subsets of variables describing a school community that would capture most of the variation between assessment results and a school's socio-economic characteristics?* The short answer is no. I attempted to use both a small set derived from EQAO's own context variables and a small set of census variables that could be easily collected at the school; in neither case did these data sets provide a strong association with assessment results.

Technical Appendix

In this technical appendix, I present the details of the methodology and results I used to arrive at the conclusions I reach in the body of Chapter 4. The appendix begins with a primer on regression methodology — the methodology I also use in Chapters 5 and 6. I then use that methodology to estimate various coefficients that describe the relationships between achievement results, dummy variables that represent the test that is written (both the year and type of test), and variables measuring the school's socio-economic context.

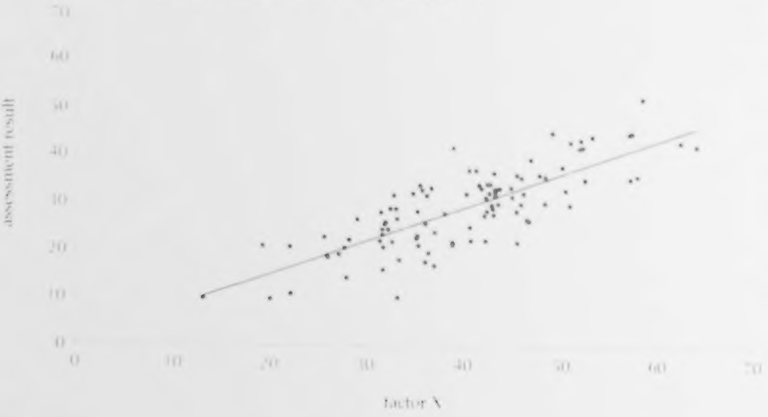
A Primer on Regression

The simplest possible relationship to estimate the relationship between the measure of school achievement and a social context variable is

$$C^1_S = \alpha + \beta_X X_S + u_S \, . \qquad \text{(A4.1)}$$

The equation has only one type of assessment result and one choice of a context variable, X_S. An estimate of the single value of β_X must emerge from the data, where each school S gets result C^1_S on the assessment and the value of X at each school S is X_S. A plot, for 80 schools, of the relationship between assessment results and X appears in Figure A4.1. The ordinary least squares estimates or regression estimates of β and β_X calculate the values that minimize the sum of squared distances (the squared values of u_S) between the actual value of the each assessment result and the corresponding point on the line drawn with intercept β and slope β_X. The positive slope of the line indicates that, on average, a school with a high value of factor X is more likely to have a higher achievement score. The precision of the estimate of the coefficient that describes that relationship depends on how close the actual observations of assessment results are to the line — that is, on the spread in the values of u_S. This metric is also related to the usefulness of knowing factor X in predicting a school's achievement score. In Figure A4.1,

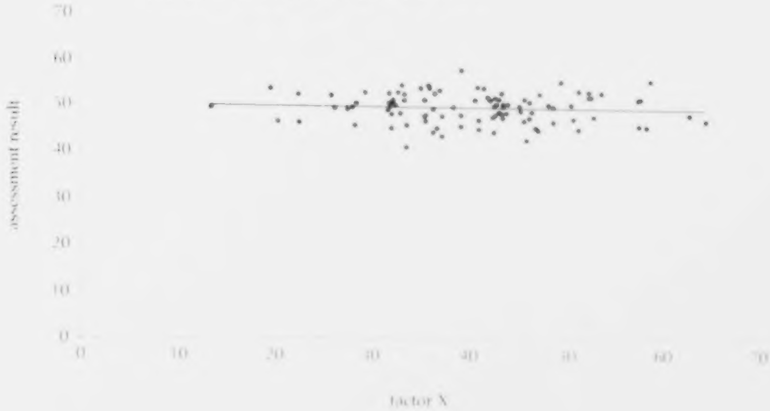
Figure A4.1: ***A Positive Relationship between Assessment Results and Factor X***



the points are relatively close to the estimated line, and the estimates of the slope of that line are precise. One can say the same thing in four ways: first, the standard error of the estimate of β_X will be small relative to its value; second, we are quite certain that the true value of β_X is not zero; third, we would reject the null hypothesis that the true value of β_X is zero and be correct in that rejection either 95 or 99 percent of the time; and fourth, if the values of u_S are small in absolute terms (that is, close to the line), the R^2 will be close to 1 (that is, high). One would be almost certain that, if a school has a high value of factor X, it will also have a high level of achievement in Figure A4.1.

In Figure A4.2, the relationship between assessment results and total enrollment at a school is negative. The slope of the estimated regression line, the value of β_X, is negative. A more careful study of Figure A4.2 reveals clearly that the value of β_X is not very precisely estimated. The average absolute values of u_S are large, and the circles that represent the actual data are not very close to the line drawn. The standard error of the estimate of β_X will be large relative to the estimate of its value. In fact, in the language of statistics, the null hypothesis that the value of β_X in Figure A4.2 is zero cannot be rejected. The R^2 in this relationship is low or close to zero.

Figure A4.2: ***A Negative Relationship between Assessment Results and Factor X***



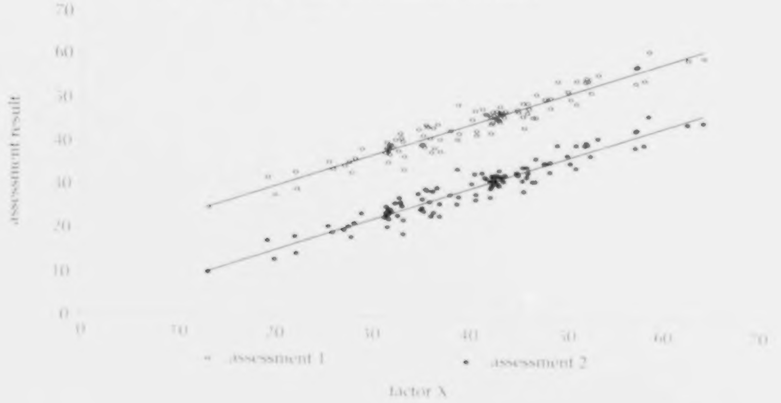
Knowing the value of factor X is not very useful in predicting the achievement level at a school in Figure A4.2.

In the body of Chapter 4, I introduced the vital concept of a dummy variable to the analysis. In this case the relationship to be estimated is

$$C^1_S = \alpha + \delta_i D_{S,i} + \beta_X X_S + u_{S,i}. \qquad \text{(A4.2)}$$

The observations of the relationship between $C^1_{S,i}$ and X_S are now available from two different assessments. These are numbered $i = 1$ and $i = 2$. There are still only 80 schools, but there are 160 dots on Figure A4.3. It is clear that these two assessments have different average achievement levels. This could be for any reason; such as an easier or harder test or improved learning. Figure A4.3 shows two quite clear lines relating the scores on assessment 1 (open circles) to factor X and the scores on assessment 2 (closed circles) to factor X. To estimate the relationship between an achievement score on either assessment 1 or assessment 2, a dummy variable takes on the value of 1 whenever an observation is drawn from an assessment of type 2 and zero whenever an assessment is drawn from type 1. The coefficient on the dummy variable is δ_i, where $i = 2$ in equation (A4.2).

Figure A4.3: ***The Relationship between Factor X and Two Assessment Results***



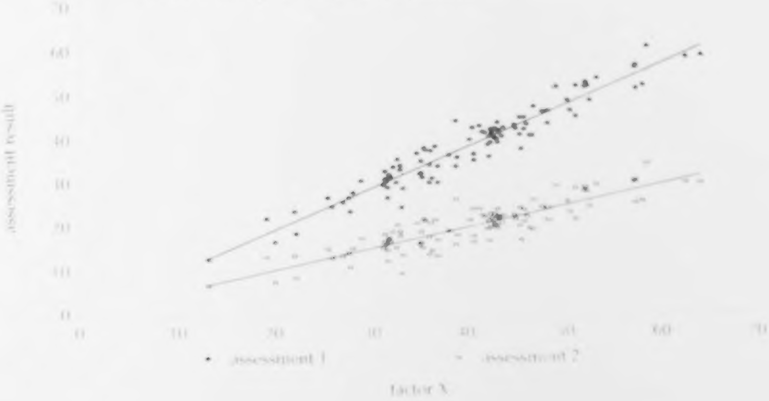
Its statistical significance depends on the way in which adding the second line reduces the absolute values of $u_{S,i}$. The value of δ_i measures how much larger (or smaller) the average achievement result is on the assessment of type 2 rather than an assessment of type 1. Notice that only one coefficient value of δ_i is estimated when there are two assessments; that single coefficient value indicates how different the average result on the second assessment is from the average result on the first assessment.

It is certainly possible that the slope relationship between factor X and the achievement score on the second assessment differs from that relationship for the first assessment. This situation is represented by the equation,

$$C^1_{S,i} = \alpha + \delta_i D_{S,i} + \beta_X X_S + \beta^X_{int} (D_{S,i} X_S) + u_{S,i}. \qquad (A4.3)$$

The coefficient β^X_{int} captures, if positive as drawn in Figure A4.4, the fact that the variable X has a stronger association with achievement scores on assessment 2 (the solid dots) than on assessment 1 (the open circles). The variable $(D_{S,i} X_S)$ is the product of the dummy variable equal to 1 for observations using assessment 2 (zero for assessment 1) and the context variable X_S. This type of variable is

Figure A4.4 *Two Assessments Where the Relationship to Factor X Differs*



called an *interaction term*, since the dummy variable interacts with the context variable. The statistical methodology tests the null hypothesis that the value of β^X_{int} is not zero. Although it is important to explore this case, for most of the variables in this analysis the value of β^X_{int} is not statistically different from zero, and simpler equations such as (A4.2) are estimated.

There is one further regression complication. We saw in the text that the equations estimated (without interaction terms) take the form of

$$C^1_{S,i,t} = \alpha + \delta_{i,t} D_{S,i,t} + \beta_X{}^1 X^1_S + \beta_X{}^2 X^2_S + u_S + u_{S,i,t}. \qquad \text{(A4.4)}$$

where there are 11 dummy variables denoted $D_{S,i,t}$ for the 11 different assessments in each grade and multiple (usually more than 2) school context variables denoted X^1_S, X^2_S, and so on. However, a key part of equation (A4.4) has yet to be discussed. In each observation of an assessment result at a school, there are two components to the error term, denoted u_S and $u_{S,i,t}$. The first component of the error term (u_S) is associated with the school. The second component ($u_{S,i,t}$) is associated with the specific grade and assessment at that school. The first component means that each assessment result at a school is not an independent observation of the effect of the socio-economic context variables on the assessment results. Instead, each school contributes one observation to its own "cluster" of data points.[18] The standard errors of the estimates of the coefficients must be adjusted for this effect.[19] Only then can the statistical significance of the estimates of β_X^1 and β_X^2, as well as any interaction terms, be appropriately calculated.

Variations in Average Achievement Results Over Time

The first equation estimated is:

$$C^1_{S,i,t} = \alpha + \delta_{1999} D_{1999} + \delta_{2000} D_{2000} + \delta_{2001} D_{2001} + u_S + u_{S,i,t}. \quad \text{(A4.5)}$$

One such equation is estimated for each type of assessment in both Grade 3 and Grade 6, for a total of six regressions. Each estimated equation contains about 3,500 schools. The coefficient on the dummy variable associated with assessments written in the 1999–2000 academic year is δ_{1999}. In Table A4.1, the estimated value of δ_{1999} for Grade 3 reading is 4.29. In 1999–2000 at the average school, the percentage of students that achieved either a Level 3 or Level 4 result on the Grade 3 reading assessment *increased* by 4.29 percentage points from the base year assessment in 1998–99.

The other two coefficients in equation (A4.5), δ_{2000} and δ_{2001}, are associated with indicator or dummy variables D_{2000} and D_{2001}, respectively, and mean that the Grade 3 reading results in 2000–01

---

[18] There are actually two clusters in the data: a school-level cluster across all assessments in all years at a school and a grade-level cluster within the school in each academic year. Only the first cluster is used to adjust the standard errors. All dependent variables are collected at the level of the school. There is no information specific to the grade at the school in different years.

[19] This issue is discussed in Deaton (1997). The estimates in this study were carried out using Stata, Version 8 (for a description of this statistical package, see www.stata.com). The estimation command in Stata is regression with the option "cluster." The correction to the standard errors is also discussed in the Stata manuals. It is an important adjustment.

Table A4.1: *Variation in Assessment Results across Years*

	Coefficients on Dummies[a] (Standard Error)			Probability Values on Tests of Equality of Coefficients[b] Null Hypothesis Tested		
	δ_{1999} 1999–2000 (1)	δ_{2000} 2000–01 (2)	δ_{2001} 2001–02 (3)	$\delta_{1999} = \delta_{2000}$ (4)	$\delta_{2000} = \delta_{2001}$ (5)	$\delta_{1999} = \delta_{2001}$ (6)
Grade 3						
Reading	4.29 (0.28)	1.96 (0.30)	5.57 (0.31)	0.27	0.00	0.00
Writing	-0.13 (0.28)	-0.60 (0.30)	1.24 (0.30)	0.10	0.00	0.00
Mathematics	1.55 (0.32)	4.09 (0.35)	2.30 (0.38)	0.00	0.00	0.04
Grade 6						
Reading	2.03 (0.28)	5.16 (0.31)	5.78 (0.32)	0.00	0.04	0.00
Writing	0.42 (0.28)	4.10 (0.31)	4.80 (0.30)	0.00	0.01	0.00
Mathematics	5.04 (0.32)	7.47 (0.35)	6.80 (0.36)	0.00	0.05	0.00

[a] All coefficients are significantly different from zero except those shaded.

[b] The probability value at which the null hypothesis holds that the coefficient in the second column differs from the coefficient in the first column.

were 3.96 percentage points higher than in 1998–99 and in 2001–02 5.97 percentage points higher than in 1998–99. As in all analysis of data, however, it is important to know the likelihood that both the change in overall results and year-over-year changes are statistically significant and not just due to random events. To ascertain what has happened, it is necessary to return to equation (A4.5) and estimate the three coefficients on the dummy variables and the correct standard errors on these coefficients. Are the changes in results on assessments of statistical significance or is a change from year to year just a result of luck, either good or bad? As I hinted earlier, this is not a straightforward statistical issue, for several reasons.

The first reason — which is almost a philosophical issue — relates to the estimation of equation (A4.5) and, indeed, to all the equations in this book. If the equations were estimated using absolutely *all* 3,963 schools that existed in the province in 2000–01, the "sample" would not be a sample at all but a census. There are no standard errors associated with measurement of the average value of a variable in a census — the value is what it is. Typically, however, the equations are estimated using a sample of between 3,200 and 3,500 schools, which, admittedly, is a very large proportion of the population of schools. When does a sample become a census of a population? There is no hard and fast answer to this question. Even the national census has some nonrespondents and lost data. The very large group of schools is quite close to a census of schools some of the time, but interpreting the data set as a census would render the concept of a hypothesis test and a standard error meaningless. We actually know the values of the coefficients in the equations that describe this population. However, the results from Grade 3 and Grade 6 students writing the assessment could be considered a sample of elementary students in these grades from all grades in the province from a given year, and the equations could be interpreted as representing a sample of grades from elementary students in schools, rather than a population of schools. In this case, the issues around statistical significance become interesting. The t-statistics and p-values then have a useful interpretation. It is so rare to have a sample that even begins to approach the population that it is difficult to know which interpretation makes more sense. In the context of clustered data, allowing the error structure to reflect the effects of clustering as in a sample makes sense to me. The interpretation that is chosen, census or sample, does not affect the choice of variables associated with higher or lower assessment results or the ability to provide a ranking that places schools in their socio-economic context.

The second statistical problem is that the school results are not a data set where all assessment results are independent — that is, unrelated to each other in a statistical sense. I mentioned this earlier, but it is worth repeating, and may be easier to understand in the context of a specific example. Each school in equation (A4.5) represents a cluster of data. When we observe that a specific school has a high value on a Grade 3 reading assessment result in one year — say, a high value of the variable $C^1_{S,i,t}$ in 1998–99 — we actually know much more. Various factors at that school would have contributed to that high result, and many of the same factors would have contributed to high scores in subsequent years. The same types of families — even the same actual families — would have sent children to that school in subsequent years. The school-level factors would not have changed from year to year. Thus, we would expect a high value of $C^1_{S,i,t}$ in an assessment at a school to be associated with high values of other assessments at that school in all the other years. A correct statistical estimate of the coefficients of equation (A4.5) takes the school effect into account.

Table A4.1 presents the estimates of the three coefficients, δ_{1999}, δ_{2000}, and δ_{2001}, and the associated estimates of the standard errors for each of the six assessments. Those with experience in the analysis of statistical data will observe that the standard errors on the coefficients are actually quite large for the number of observations. It is quite difficult to estimate tightly the values of δ_{1999}, δ_{2000}, and δ_{2001} in equation (A4.5). This is the effect of clustering. Tests of a variety of null hypotheses are implicit or explicit in Table A4.1. A null hypothesis asks whether one can reject the validity of a specific numerical statement and be correct in doing so either 95 percent of the time or 99 percent of the time, depending on the level of significance of the test chosen. Tests of three null hypotheses are implicitly presented in listing the coefficients and the standard errors on the dummy variables for each academic year other than 1998–99. The coefficients in the first three columns of the table measure directly whether the assessment results in 1999–2000, 2000–01, and 2001–02 differ from the 1998–99 results. When the coefficients on the dummy variables are positive and more than twice their standard error — the values in parentheses — then the assessment results in those years are significantly higher than in the base year. If we are confident at a 95 percent confidence interval that the coefficient is statistically different from zero, there is no shading in the table.[20] In the fourth column, a null hypothesis is posed: do the results of the 2000–01 and 1999–2000 assessments differ equally from those of 1998–99? The fifth column asks whether the results of the 2001–02 and 2000–01 assessments differ equally from those of 1998–99, and the sixth column asks whether the results of the 2001–02 and 1999–2000 assessments differ equally from those of 1998–99. The fourth, fifth, and sixth columns report probability values at which the null hypothesis would be rejected. A value greater than 0.05 indicates that, with a 95 percent likelihood of being correct, one is not able to reject the null hypothesis.

What can we learn from Table A4.1? Because the coefficients on the dummy variables grouping results by year are positive and significant on 16 of 18 assessments of the same type (so that, in the same grade, reading is compared with reading, writing with writing, and mathematics with mathematics), it is possible to conclude that, after 1998–99, average results were generally higher. Neither negative coefficient — the estimated values are $\delta_{1999} = -0.13$ and $\delta_{2000} = -0.60$ — is actually statistically different from zero at conventional levels. Both estimates are small, both in an absolute sense and relative to other estimated values of δ. The positive estimates of the various other values of δ on the reading and mathematics assessments are large and much different from zero in both Grade 3 and Grade 6. The ability to master the curriculum as measured by the assessment has increased since the 1998–99 assessment.

It is disturbing that there are two exceptions to the steady increase in assessment results. These reductions are shown because the δ_{2001} coefficient value, 2.30, is significantly less than the δ_{2000} coefficient value of 4.09 on the Grade 3 mathematics assessment. The probability value on rejection of the null hypothesis that $\delta_{2001} = \delta_{2000}$ is 0.00.

---

[20] I realize it is an oddity that only the coefficients that are not statistically significant are shaded in Table A4.1 and subsequent appendix tables. In this analysis, however, there are many fewer insignificant coefficients than coefficients significantly different from zero. This is a function of both the large sample and the strong relationships between socio-economic context variables and assessment results. It seems easier to have the reader assume all relationships are statistically significant unless otherwise identified.

The same statement can be made when comparing the 2000–01 and 2001–02 Grade 6 mathematics outcomes. How can one explain the reduction of test results for the entire province in one year? Remember that labelling higher average scores as an improvement in learning outcomes makes sense only if one is absolutely certain that the assessment instrument and its grading are consistent over the years. What does a decline in the average score then mean? As already noted in the body of the chapter, as a professor who has created several hundred examinations over the years, my personal belief is that the likelihood of absolutely comparable assessments is quite low. It is fairly straightforward to think of a situation in which an assessment was accidentally increased in difficulty or marked in a different way. It is very hard to think of a situation in which very large samples of similar children and teachers suddenly reduce their performance from one academic year to the next. For this reason alone, I find that using the assessment results to evaluate relative school success, the exercise undertaken in Chapter 6, is the more promising use of the Ontario assessment process. As I argued in the body of the text, the assessments needed to ask if Ontario students are improving are the multi-jurisdictional assessments administered by EQAO.

An equation very similar to (A4.5) is

$$C^1_{S,i,t} = \alpha + \hat{\delta}_{\mathrm{write}} D_{\mathrm{write}} + \hat{\delta}_{\mathrm{math}} D_{\mathrm{math}} + u_S + u_{S,i,t}. \tag{A4.6}$$

This equation is estimated separately for each year of data for the 3,500 schools. The coefficient estimates of $\hat{\delta}_{\mathrm{write}}$ and $\hat{\delta}_{\mathrm{math}}$ are found in Table A4.2. The base test in each year is the reading test. Table A4.2 makes it very clear that, within a year, the assessments in reading, writing, and mathematics typically have different average values of the percentage of students that attains Level 3 or Level 4. Grade 3 students, in every year, were more likely to achieve a satisfactory level of achievement in writing and mathematics than in reading. By Grade 6, these differences appear to be reversed and are much smaller in magnitude. The differences remain statistically significant in all but one case. Again, the differences are difficult to interpret.

Table A4.2: *Variation in Assessment Results across Tests*

	Coefficients on Writing Dummy[a]		Coefficients on Mathematics Dummy	
Grade 3				
1998–99	7.10	(0.21)	10.55	(0.21)
1999–2000	2.67	(0.18)	7.81	(0.21)
2000–01	2.53	(0.18)	10.69	(0.20)
2001–02	7.77	(0.20)	7.28	(0.20)
Grade 6				
1998–99	-0.90	(0.20)	-2.15	(0.20)
1999–2000	2.52	(0.19)	0.85	(0.19)
2000–01	-1.96	(0.19)	0.14	(0.21)
2001–02	1.88	(0.18)	1.13	(0.21)

[a] Standard errors are in parentheses. All coefficients are significantly different from zero except as shaded.

If the assessment instrument were perfect and correctly measured the ability of students to master the curriculum, the coefficients on these dummy variables state that, in all years, Grade 3 students in Ontario are relatively better at writing and mathematics than at reading. By Grade 6, these differences are much smaller and in one case not significant. For the purpose of understanding the effect of socio-economic factors on achievement results, Tables A4.1 and A4.2 tell us it is important to allow the average score on an assessment to vary by type of assessment as well as by the year of the assessment, then measure the size of the association of socio-economic factors in the school community on school results. This is our next step.

Identifying School Community Context Variables Associated with Higher or Lower Test Scores

The equations estimated to identify the relationships between context variables and assessment results take the form of

$$C^{1}_{S,i,t} = \alpha + \delta_{\text{test}} D_{\text{test}} + \beta_X^{1} X^{1}_{S} + \beta_X^{2} X^{2}_{S} + u_S + u_{S,i,t} \,. \qquad (A4.7)$$

and

$$\begin{aligned} C^1_{S,i,t} = {} & \alpha + \delta_{test} D_{test} + \beta_X{}^1 X^1{}_S + \beta^1{}_{int} (D_{test} X^1{}_S) \\ & + \beta_X{}^2 X^2{}_S + \beta^2{}_{int} (D_{test} X^2{}_S) + u_S + u_{S,i,t} \,. \end{aligned} \tag{A4.8}$$

There are more than two context variables and more than one dummy variable in the actual estimated equations, so a simplified notation is used here.

The constant terms in these equations denoted α do not have a simple interpretation.[21] The coefficients denoted δ_{test} on the 11 dummy variables in each equation do have a clear interpretation. A dummy variable is equal to 1 for an observation from a specific type of assessment, whether reading, writing, or mathematics, in one of the four academic years. There is thus a total of 12 assessments in one estimating equation. These coefficients measure the change in the percentage of students achieving at Level 3 or 4 on the assessment that the dummy variable represents relative to the assessment without a dummy variable. The results in Tables A4.1 and A4.2 mean that it is very important to allow the different types of assessments in different years to have different average results. These effects are included "in passing" in estimating equations (A4.7) and (A4.8), because they are not of central interest in the analysis of the measurement of the association between school assessment results and the variables that describe the characteristics of the school community. To avoid further clutter, the estimates of the values of the coefficients on these dummy variables are not reported in the tables.

The coefficients on the dummy variables are always statistically significant. The coefficients on the interaction terms, $\beta^1{}_{int}$ and $\beta^2{}_{int}$, the product of the dummy variables and X^1 and X^2 in equation (A4.8), allow the relationships between the assessment results and the school context variables to vary by type of assessment. The interaction coefficients are tested for their statistical significance and the results are presented at various places in the analysis.

---

[21] The constant term is the average achievement on the base year and base type of test adjusted for the values of the community context variables at the average school. The constant term no longer has a useful direct interpretation.

Tables A4.3, A4.4, and A4.5 present the detailed estimates of the coefficients on different relationships between various combinations of the school community variables and the percentage of students achieving at Level 3 or 4 using variants of equations (A4.7) and (A4.8). Table A4.3 presents estimates from the Grade 3 assessments, Table A4.4 presents estimates of equations from the Grade 6 assessments, and Table A4.5 combines information on the two assessments.

The general format of the estimated equations follows either equation (A4.7) or (A4.8) but uses more than two school community context variables. When an equation similar to (A4.7) or (A4.8) is estimated with only Grade 3 assessments, there are 11 dummy variables forming groups for 11 of the 12 assessments (the reading assessment in 1998–99 is represented by the omitted dummy variable). When the equation is estimated using the Grade 6 assessment results, there are also 11 dummy variables whose coefficients capture differences in the average level of achievement by assessment and year in Grade 6 (the reading assessment in 1998–99 is represented by the omitted dummy variable). In Table A4.5, when the effect of socio-economic school community variables on achievement is measured using both Grade 3 and Grade 6 assessments, there are 23 assessment-related dummy variables, and the Grade 3 reading assessment in 1998–99 is represented by the omitted dummy variable. The coefficients on these variables are always statistically significant but are not presented in Tables 4.3, 4.4, or 4.5.

In this part of the analysis, three questions of direct interest are posed. First, how much of the variation in achievement results in a given year can be explained by variations in the socio-economic variables that describe a school community? The answer to this question is discovered by looking at the R^2 that appears in the last line of the three tables. Remember that when the R^2 is unity or close to unity, all or nearly all of the variations in assessment results would be associated with variations in the school's socio-economic context. In the single variable case of Figure A4.1 with an R^2 of unity, all the observations would fall exactly on the regression line and the deviation from the regression line would be zero. This is simply not the case for any version of any equation estimated. This

Table A4.3: *The Association of Socio-Economic School Community Variables with Grade 3 Assessment Scores*[a,b]

			All Variables with Interaction Dummies, by Type of Test (Standard Error)			
Variable (1)	Mean (Standard Deviation) (2)	Base Model All Variables (Standard Error) (3)	Base Coefficient (4)	Interaction with Writing (5)	Interaction with Math (6)	Scaling Calculation[c] (7)
Single-parent families (%)	22.0 (8.6)	-0.11 (0.04)	-0.11 (0.04)	-0.05 (0.03)	0.06 (0.03)	-0.95
English or French as home language (%)	91.5 (9.5)	-0.10 (0.04)	-0.04 (0.05)	-0.026 (0.03)	-0.14 (0.03)	-0.95
Aboriginal status (%)	1.62 (2.6)	-0.55 (0.08)	-0.50 (0.08)	0.06 (0.05)	-0.18 (0.06)	-1.49
Recent immigrants (in Canada < 5 years) (%)	4.3 (5.6)	0.13 (0.08)	0.08 (0.08)	-0.029 (0.05)	0.16 (0.06)	0.73
Born outside Canada, U.S., or U.K. (%)	19.1 (17.0)	0.04 (0.03)	0.021 (0.03)	0.07 (0.02)	-0.01 (0.02)	0.68
Living in detached home (%)	65.7 (23.1)	0.06 (0.01)	0.05 (0.01)	0.01 (0.01)	0.03 (0.01)	1.39
Log of income[d]	10.9 (0.28)	7.05 (1.28)	7.20 (1.36)	-0.34 (0.37)	-0.11 (0.98)	1.97
Moved in past year (%)	13.4 (4.0)	-0.39 (0.10)	-0.40 (0.10)	-0.04 (0.06)	0.06 (0.07)	-1.56

Moved in past 5 years (%)	41.1 (9.4)	-0.07 (0.03)	-0.08 (0.04)	-0.065 (0.02)	-0.05 (0.02)	-0.66
Unemployment rate (%)	6.5 (3.4)	-0.29 (0.10)	-0.23 (0.11)	-0.006 (0.06)	-0.15 (0.07)	-0.99
Adults without secondary school diploma (%)	45.1 (9.7)	-0.22 (0.04)	-0.20 (0.05)	-0.001 (0.03)	-0.05 (0.04)	-2.13
Adults with some university education (%)	21.9 (11.1)	0.23 (0.04)	0.28 (0.05)	-0.05 (0.03)	-0.08 (0.03)	2.55
Total enrollment	350 (165)	0.003 (0.003)	0.005 (0.003)	0.0001 (0.002)	0.0019 (0.002)	0.50
Total enrollment squared	150,329 (151,434)	-0.000005 (0.000003)	-0.000003 (0.000003)	0.000001 (0.000001)	-0.000006 (0.000002)	-0.91
Sample size	41,122	41,111	41,111			
Number schools		3,584	3,584			
R^2		0.264	0.266			

---

[a] All coefficients are significantly different from zero except those shaded.

[b] The average value of the dependent variable, the percentage of students at Level 3 or Level 4 is 51.7 percent (standard deviation 18.9).

[c] The scaling calculation multiplies the estimated coefficient on a variable by a one standard deviation increase in that variable.

[d] The average value of the level of income is $62,222 (in 1995 dollars) (standard deviation 18,820).

Table A4.4: *The Association of Socio-Economic School Community Variables with Grade 6 Assessment Scores*[a,b]

Variable *(1)*	Mean (Standard Deviation) *(2)*	Base Model, All Variables (Standard Error) *(3)*	All Variables with Interaction Dummies, by Type of Test (Standard Error) Base Coefficient *(4)*	Interaction with Writing *(5)*	Interaction with Math *(6)*	Scaling Calculation[c] *(7)*
Single-parent families (%)	21.7 (8.4)	–0.15 (0.04)	–0.14 (0.04)	–0.04 (0.02)	–0.001 (0.03)	–1.26
English or French as home language (%)	91.8 (9.3)	–0.05 (0.05)	–0.03 (0.03)	–0.04 (0.03)	–0.18 (0.03)	–0.47
Aboriginal status (%)	1.6 (2.6)	–0.48 (0.10)	–0.50 (0.10)	0.07 (0.05)	–0.05 (0.06)	–1.25
Recent immigrants (in Canada < 5 years) (%)	4.0 (5.2)	0.13 (0.08)	0.05 (0.09)	–0.04 (0.04)	0.38 (0.06)	0.67
Born outside Canada, U.S., or U.K. (%)	18.3 (16.6)	0.001 (0.03)	0.04 (0.04)	0.008 (0.02)	–0.12 (0.02)	0.02
Living in detached home (%)	66.8 (22.4)	0.034 (0.01)	0.04 (0.01)	–0.03 (0.009)	0.04 (0.01)	0.76
Log of income[d]	10.99 (0.28)	7.53 (1.26)	6.42 (1.28)	0.94 (0.91)	2.37 (0.88)	2.11
Moved in past year (%)	13.3 (4.0)	–0.20 (0.10)	–0.23 (0.10)	–0.01 (0.06)	0.10 (0.07)	–0.80

Moved in past 5 years (%)	40.1 (9.3)	–0.07 (0.04)	–0.06 (0.04)	0.01 (0.03)	0.10 (0.07)	–0.65
Unemployment rate (%)	6.4 (3.3)	–0.33 (0.1)	–0.33 (0.1)	0.12 (0.06)	–0.02 (0.07)	–1.09
Adults without secondary school diploma (%)	45.0 (9.6)	0.09 (0.05)	–0.07 (0.05)	–0.02 (0.03)	0.04 (0.04)	0.86
Adults with some university education (%)	21.9 (11.1)	0.51 (0.05)	0.51 (0.04)	–0.07 (0.03)	0.004 (0.03)	5.66
Total enrollment	353 (165)	–0.001 (0.003)	0.01 (0.003)	0.004 (0.001)	–0.003 (0.002)	–0.17
Total enrollment squared	152,475 (150,359)	0.000009 (0.000002)	0.000009 0.000003	0.000002 (0.000001)	0.000002 (0.000002)	1.35
Sample size	38,769	38,762	38,762			
Number schools		3,403	3,403			
R^2		0.28	0.29			

---

[a] All coefficients are significantly different from zero except those shaded.

[b] The average value of the dependent variable, the percentage of students at Level 3 or Level 4 is 51.1 percent (standard deviation 18.1).

[c] The scaling calculation multiplies the estimated coefficient on a variable by a one standard deviation increase in that variable.

[d] The average value of the level of income is $66,296 (in 1995 dollars) (standard deviation 18,803).

Table A4.5: *An Interaction Model of Achievement in Grade 3 and Grade 6*

Variable	Base Coefficient (Standard Error)	Interaction with Grade 6 (Standard Error)
Single-parent families (%)	−0.10 (0.04)	−0.05 (0.05)
English or French as home language (%)	−0.10 (0.05)	0.04 (0.05)
Aboriginal status (%)	−0.54 (0.08)	0.05 (0.09)
Recent immigrants (in Canada < 5 years) (%)	0.14 (0.08)	−0.05 (0.08)
Born outside Canada, U.S., or U.K. (%)	0.04 (0.04)	−0.03 (0.03)
Living in detached home (%)	0.06 (0.01)	−0.01 (0.0)
Log of income	7.53 (1.11)	−0.49 (0.59)
Moved in past year (%)	−0.39 (0.10)	0.18 (0.11)
Moved in past 5 years (%)	−0.07 (0.01)	−0.0007 (0.04)
Unemployment rate (%)	−0.28 (0.10)	0.004 (0.011)
Adults without secondary school diploma (%)	−0.21 (0.05)	0.12 (0.05)
Adults with some university education (%)	0.23 −0.04	0.25 (0.04)
Total enrollment	−0.003 (0.003)	−0.018 (0.003)
Total enrollment squared	−0.000005 (0.00003)	0.000014 (0.000006)
Number of observations	79,873	
Number of schools	3,774	
R^2	0.27	

Note: Shading indicates the coefficient on that variable is not statistically different from zero.

is an important result. Other factors are associated with higher or lower results and are not included in the context variables that describe the school's socio-economic characteristics. The highest values of the R^2 indicate that about 25 percent of the variation in achievement results over all assessments over all years is associated with variations in the socio-economic variables as well as variations in the average assessment results, while, as noted earlier, fully 75 percent of the variation in assessment results in a given year on a given assessment is not explained. It is clear even at this stage of the analysis that school results vary enormously even among *very similar schools* and from one year to another at the same school. It is simply not true that assessment results depend *only* on school context.

The second question is: beyond the dummy variables that group assessments by year and type, which socio-economic context variables that describe a school community are most closely related to variations in school assessment results? The variables that represent these factors are listed in the first column of Tables A4.3 and A4.4. The second column presents the average value of each row variable and its standard deviation. The third column provides estimates of the coefficient on the row variable in the base model specification without interaction effects. In both these specifications, the estimated coefficient is the effect of that row variable on assessment results averaged over the three assessments (reading, writing, and mathematics).

As an example, the first estimated coefficient (the coefficient on the single-parent measure) in Table A4.3 has a value of –0.11 and a standard error of 0.04. Because the coefficient is more than twice its standard error, it is very likely that this coefficient is actually different from zero and negative. Thus, in school communities with a higher percentage of single parents, *all other factors being the same*, achievement results on all Grade 3 assessments (reading writing, and mathematics) are usually lower. For each percentage point increase in the percentage of single parents in the school community, *all other factors being the same*, it is predicted that 0.11 of a percentage point fewer students will achieve either Level 3 or Level 4. One cannot say why this association occurs, only that the association of large numbers of single parents and lower results is a feature of the data.

The last column of Tables A4.3 and A4.4 helps the reader to scale the coefficient estimate in a different and perhaps more useful way. It takes the estimated coefficient and calculates the size of the predicted effect on school achievement levels of a one standard deviation change in the value of that school community characteristic.

In the example above, 22 percent of families in the average elementary school community are headed by single parents. The standard deviation of the single-parent school community variable is 8.6 percentage points in the Grade 3 sample. So, at a school with 8.6 percentage points more single parents than another school with the average percentage of single parents, the assessment score would be predicted to fall by 0.95 of a percentage point, *all other school context variables being the same.*

The remainder of the values in the last column of Tables A4.3 and A4.4 are read in the same way. The scaling calculations use the estimates of the coefficients on the base model without the interaction effects found in the second column. The relationships revealed by the base model are discussed in the body of Chapter 4.

A slightly technical issue arises concerning the estimation of the effects of income on achievement results. The income variable in the estimated relationships is actually the natural logarithm of income. For reference, the average level of household income, a more intuitive concept, is found in the notes to Tables A4.3 and A4.4. Household income is highly skewed to the right (see Chapter 3), so that there are a few schools with very high average household incomes. In economics, it is common to think of the association of income with an outcome as nonlinear. When the logarithm of income is entered in the equation, equal percentage increases of income have the same association with assessment results. To understand the effect of income on school results, suppose a school community increases its average household income from $60,000 to an enormously high $120,000. In this example, the log of income variable would increase by 0.69 units and the predicted achievement level would rise by 4.9 percentage points. This is a very large effect.

However, 2.85 percentage points of this effect — almost 60 percent of the total effect — occur as average household income rises from \$60,000 to \$90,000. Using the logarithm of income as the X variable picks up this effect.

The third question the methodology needs to address is the following: are the coefficients on the various interaction coefficients statistically significant — that is, do the socio-economic variables have similar relationships to different types of assessments (reading, writing, and mathematics) or to assessments written in different grades (Grade 3 and Grade 6)?

The answer to this question is found in the coefficient estimates in Tables A4.3 and A4.4 under the heading "All Variables with Interaction Dummies, by Type of Test." Earlier, we considered the possibility of interaction terms where the relationships between context variables and reading, writing, and mathematics assessment results need not be the same. These hypotheses can be tested by estimating versions of equation (A4.8) with all socio-economic variables included and asking whether the coefficients on the interaction variables, denoted β^1_{int} and β^2_{int} in equation (A4.8), differ from zero.

As in the case of the other coefficients on indicator or dummy variables, the change in the effect of variable X on the specific test is measured relative to a base case. As an example, consider the row of coefficients on the percentage of single parents in the school community. When the base model is estimated under the heading "All Variables with Interaction Dummies, by Type of Test," the coefficient on the single-parent variable is –0.11. This is the effect of an additional percentage of single parents on the reading assessment result, the base assessment. The association of single parents with lower assessment results is more negative for the writing assessment (the point estimate of the coefficient on the interaction variable is –0.05, so the total predicted effect on a writing assessment is –0.16) and less negative on the mathematics assessment (the point estimate of the coefficient on the interaction variable is 0.06 and the total effect is –0.05). The coefficient on the mathematics-single-parent interaction variable is of statistical significance and is not shaded.

The results include a variety of interaction effects, some more interesting than others. Again, I refer the reader to the main body of this chapter for a summary of these associations. In most cases, however, the coefficient on the interaction term is small relative to the base coefficient even when of statistical significance. Are the interaction variables worth adding to the explanatory equation? The increase in the explanatory power of the equation when these variables are added is tiny, so the rest of the book ignores these interaction effects.

The second possible interaction is the difference between Grade 3 and Grade 6 assessments and their relationship to socio-economic variables. This is dealt with in two ways, as shown in Tables A4.4 and A4.5. Table A4.4 presents the same set of estimates of coefficients based on Grade 6 assessment results. There are some differences and some similarities between results from the two grades, which have already been discussed in the body of the chapter. The coefficient on the single-parent variable is negative. The coefficient on the variable that measures the percentage of the school community that speaks an official language at home is not of statistical significance except as an interaction variable reducing mathematics scores. Schools with larger populations drawn from recent immigrants substantially outperform other schools in the Grade 6 mathematics assessment, but the effect of longer-term immigrants on Grade 6 results is quite weak. A large proportion of aboriginals in a school community lowers average scores, and the effect is substantial. The economic variables have strong associations with Grade 6 assessment results. Higher average incomes in the school community are strongly associated with higher scores; higher unemployment rates are associated with lower scores; more detached homes in the school community are associated with higher scores. The one-year mobility variable is associated with lower scores, as it was in Grade 3. The variable measuring five-year mobility in the school community does not have a statistically significant association with Grade 6 scores. This is a change from the Grade 3 analysis. There is also a very interesting and quite dramatic change from Grade 3 to Grade 6 in the association of parental education and assessment results. By Grade 6, an increase in the percentage of the school population without a high school diploma is no longer associated with lower or higher achievement. But the percentage of the adult population in the school community with some university education has a very strong association with higher Grade 6 achievement results. The estimate of the coefficient on this variable is double that of the Grade 3 estimate.

Table A4.5 presents results that include a formal test of the equality of coefficients on each variable that describes the school community in its association with higher or lower Grade 3 and Grade 6 assessment results. The first column of coefficients presents estimates of the association between the percentage of students in Grade 3 achieving at Level 3 or Level 4 and the various variables that describe school communities. As must be the case, these coefficients are very similar to the coefficients in the second column of Table A4.3. Grade 3 is the base of this model, while the Grade 6 observations provide the interaction variables, as in equation (A4.7).[22] A dummy variable is created equal to 1 when an assessment is from Grade 6. Then, an interaction variable similar to those created earlier is the product of this dummy variable and the school community characteristic in question. The coefficient on each interaction variable is reported in the second column of Table A4.5. If the coefficient on a specific interaction variable differs significantly from zero, that school community variable is associated with higher or lower Grade 3 assessment results in a different way than the Grade 6 assessment results.

---

[22] The only differences between the coefficients in the base model in Table A4.3 and the first column of Table A4.5 arise from variations in the sample of schools used. Estimation of the interaction effects requires a school to have observation from both Grade 3 and Grade 6, but not all schools contain both grades. For the same reason, the standard deviations of the variables differ slightly across the two grades.

Notice that all but three of the coefficients on interaction variables are not significantly different from zero. There is no difference in the association of the language, aboriginal, immigration, housing, mobility, and unemployment rate variables. There is a small difference in the effect of the square of enrollment — larger schools, as measured by this variable, are associated with higher Grade 6 assessment results. There are, however, large differences in the coefficients on the two important variables that measure the educational characteristics of the school community. As already noted, a large percentage of adults in the school community with less than a high school education lowers Grade 3 results. The coefficient on the associated interaction variable shows that this measure of education no longer has a strong negative association with Grade 6 results — the sum of the effect is –0.21 + 0.12, or –0.09. However, an increase in the percentage of people with a university degree in the school community is associated with higher results on the Grade 6 assessments by much a much larger amount than on those of Grade 3. The total Grade 6 coefficient is twice the size of the Grade 3 coefficient. These coefficient differences mean that, in the rest of the book, it is appropriate to estimate separate equations and separate coefficients for the relationship between the school community variables and the results on the Grade 3 and Grade 6 assessments.

Formal Estimates of the Relationships between Test Results and Two Small Subsets of Socio-Economic Variables

In the body of this chapter, I discussed the possibility that two small subsets of "cheap and easy" socio-economic variables might do all the work needed to explain variations in assessment results. This part of the technical appendix presents the regressions using those subsets.

In its web site postings of school assessment results for the 2000–01 and 2001–02 academic years, EQAO provides four context variables: the percentage of students enrolled in ESL programs, the percentage receiving special education support; the percentage that

Table A4.6: ***EQAO Context Variables as Predictors of Results***

	Grade 3 (Standard Error)	Grade 6 (Standard Error)
Students in an ESL program (%)	–0.31 (0.10)	–0.18 (0.09)
Students using the language of instruction as their home language (%)	–0.09 (0.06)	–0.12 (0.05)
Students in special education (%)	–0.66 (0.11)	–0.80 (0.10)
Number of assessments	4,130	4,128
Number of schools	352	355
R^2	0.11	0.13

Note: Shading indicates the coefficient on that variable is not statistically different from zero.

speaks only or mostly a language other than English at home, and the percentage that speaks a language other than English as often as English at home.

From EQAO's alphabetical list of schools on the web site, I took every tenth school to create a random sample of 350. I then estimated equation (A4.6) for the 350 schools and the EQAO variables. The estimated coefficients on these variables are presented in Table A4.6. The language variable used — the percentage of students that speaks the language of instruction at home — has a negative sign and is statistically significant in the Grade 6 equation. This is consistent with the counterintuitive result that a large proportion of immigrants is associated with higher scores after the language effect has been taken into account. The language effect is directly measured in the EQAO variables as the percentage of students in an ESL program. As this percentage increases, assessment results tend to fall. Then, controlling for that effect, assessment results tend to rise as languages other than English are spoken at home. If a class contains more special needs students, this is associated with lower assessment results, which makes sense, since Category One is the percentage of *all* students in Level 3 or Level 4. If many or

Table A4.7: ***The Association of Assessment Results with a Small Subset of Census Variables***

Variable	Grade 3 Coefficient (Standard Error)	Grade 6 Coefficient (Standard Error)
English or French spoken at home (%)	–0.19 (0.03)	–0.15 (0.05)
Living in detached home (%)	0.05 (0.01)	0.03 (0.01)
Moved in the past year (%)	–0.93 (0.07)	–0.80 (0.08)
Total enrollment	0.031 (0.004)	0.01 (0.004)
Total enrollment squared	-0.000002 (0.000004	-0.000008 (0.000004)
Sample size	41,122	38,769
Number of schools	3,585	3,404
R^2	0.11	0.07

Note: Shading indicates the coefficient on that variable is not statistically different from zero.

most special needs students are exempted, then the percentage of *all* students that achieves Level 3 or Level 4 as a proportion of the class must fall. If many or most ESL students are exempted, then the percentage of *all* students that achieves Level 3 or Level 4 as a proportion of the class must fall as the proportion of ESL students rises.

The most relevant part of Table A4.6 is the large reduction in the R^2 values from those in most estimates of similar equations presented in Tables A4.3, A4.4, and A4.5. These estimates show that, even with the advantage of data that directly describe the students writing the assessment rather than data related to characteristics of the whole school community, the EQAO context variables are not as useful as the variables created from the postal codes and census. The equations that use only the EQAO variables simply do not explain as much of the variation in assessment results as the census-based equations.

In the body of the chapter, I also noted that it would be easy for schools to collect "cheap and easy" variables about individual students: a one-year mobility measure, a language variable, and a measure of the percentage of the school community that lives in detached housing, and two enrollment variables. I use all five of these variables to estimate a version of equation (A4.7), and present the results in Table A4.7. If these five variables were able to account for the variations in school achievement results, it would not be necessary to examine the more extensive socio-economic data provided by the census; unfortunately, however, this is not the case. The R^2 falls by a large amount when only these five variables are used for either Grade 3 or Grade 6. If it is useful to continue with a small number of context variables, it would have to be done at the level of the class or the individual. This experiment may be worth retrying with a sample of students from classes at a future date.

5

An Analysis of the Exemption Process

Some Ontario students are exempted from the Grade 3 and Grade 6 assessments undertaken by the Education Quality and Accountability Office (EQAO). These exemptions are not granted lightly. As EQAO states:

> Teachers and principals must make every effort to enable students with special needs to participate with their peers in all aspects of the assessment....The principal is authorized to exempt a student from part or all of an assessment in a situation where, even with all possible accommodations, the student would be unable to participate productively and/or where the student's participation would be harmful. Before exempting a student, the principal is required to consult with the student's teacher and obtain written parental consent.[1]

Indeed, as Table 4.1 showed, the exemption rate is quite low, with average rates across schools varying by assessment and year between 3.85 percent and 6.90 percent. It is, however, extremely difficult to apply either the EQAO exemption criteria quoted above or any other criteria that one could imagine in an objective and consistent way. This is simply an observation, not a criticism.

---

[1] See EQAO's web site: www.eqao.com/eqao/home_page/o5e.html.

This chapter presents a brief analysis of the data that describe the process by which students are exempted from assessments. There are two reasons to be interested in exemption rates. First, an analysis of the exemption data is also an indirect examination of the process that administers the assessments. If there is no pattern in exemptions across school boards over time, it seems more likely that the assessments are being administered consistently across the province. The EQAO criteria indicate no reason to expect systematic variations in the exemption process across school boards. In any case, this variation, though statistically significant, is not of much practical significance.

The second and more important reason to undertake an analysis of the exemption process is that some analyses of assessment results use EQAO's "Method 2,"[2] whereby the results are presented as the percentage of students achieving Level 1, 2, 3, or 4 calculated as a *percentage of the total number of students who actually provided sufficient material to be assessed*. By definition, therefore, exempted students are excluded from this group. Because the benchmark groups of students who are usually considered to meet provincial education standards are those performing at Levels 3 and 4, it thus becomes completely unclear what Method 2 results mean under an exemption process that has any hint of inconsistency. If the exemption process were applied consistently across all boards and years — a hypothesis rejected in this chapter — then Method 2 results might be considered a way to measure teaching success from the subset of students with the *potential* to master the curriculum. The difficulty of developing a method of assessing that potential is, however, obvious.

In fact, the main reason to avoid using Method 2 results is that they are open to manipulation. If a school board or even individual school felt that achieving a high Method 2 result were important, then that board or school might be tempted simply to exempt from the assessment as many students as possible who were expected to achieve only Level 1 or 2. In the extreme case, a school could have 100 percent of the students who actually write a test achieve Level 3 or 4 simply by exempting all those who would have scored lower.

---

[2] For example, the Fraser Institute's evaluation of Ontario's elementary schools uses Method 2 results (see Cowley and Easton 2003). Other aspects of the Fraser Institute's report card are considered in Chapter 6.

Looking at the statistical properties of the exemption process gives some insight into whether such manipulation is already occurring. Given the small amount of evidence of inconsistent application of the exemption process, it is more appropriate to use assessment results that include *all* students in any comparison of schools (EQAO's "Method 1"). Indeed, EQAO itself suggests that Method 1 is more reflective of its mandate to help the education system meet the needs of all children in Ontario. The Ontario government has also made a commitment to improving assessment results;[3] importantly, however, it has not committed to using Method 1, rather than Method 2, as a way of measuring that improvement. In my view, it would be more helpful if EQAO and the boards simply stopped producing and presenting Method 2 results. Having two sets of results confuses the issue.

A final observation is that, since it is more difficult to link exemption rates to socio-economic variables than it is to link achievement results to those variables, it is also more difficult to use statistical techniques to identify schools or boards where exemption rates are either too high or too low. This may be a very practical reason to avoid the use of Method 2 results.

---

[3] In October 2003, responding to the release of EQAO test results, Ontario Minister of Education Gerard Kennedy stated that he would work with educators, school boards, and parents in a new spirit of cooperation to improve student achievement:

> We recognize that our kids need strong reading, writing, and math skills to achieve success in school and after they graduate… Clearly these results show that we are not making as much progress as everyone would like….We are committed to finding ways to improve student performance.

See Ontario Ministry of Education, *News Release*, October 30, 2003; available at web site: http://ogov.newswire.ca/ontario/GPOE/2003/10/30/c6759. html?lmatch=&lang=_e.html.

A Broad Picture of Exemptions

A broad picture of the exemption process is painted in Table 5.1 and Figure 5.1. The table presents average exemption rates, calculated across schools, on 24 assessments over four academic years — that is, the reading, writing, and mathematics assessments for Grades 3 and 6 for 1998–99 through 2001–02. Using the same techniques described in Chapter 4, I tested and rejected the specific statistical hypothesis that exemption rates are equal across years on the same assessment and in a single year on different assessments.

The differences in exemption rates from one year to the next are statistically significant but not of much practical significance. The highest exemption rate is 6.9 percent, in the Grade 3 reading assessment in 1999–2000, while the lowest rate is 3.84 percent, in the Grade 6 mathematics assessment in 2000–01. It is interesting that the exemption rates in Grade 6 are considerably lower than those in Grade 3. Recall that these are very large groups of both schools and students. It seems very unlikely that the characteristics of the entire pool of Grade 3 or Grade 6 students would actually have changed very much from year to year. It seems much more likely that the exemption criteria were applied differently over time, as common sense suggests would be the case. In turn, this suggests Method 2 presentations of results are not exactly comparable across years.

More mysteriously, Table 5.1 and Figure 5.1 indicate that exemption rates vary by type of assessment in Grade 3 but not in Grade 6. Since the exemption criteria themselves offer no obvious reason for this variation, an outsider to the exemption process, such as this writer, might presume that the major exemption issue is language: clearly, there are more language exemptions than mathematics exemptions in Grade 3 but not in Grade 6.

Despite the statistical significance of varying exemption rates over time, is this variation actually important to the interpretation of the elementary school assessment process? The answer likely is no. If the primary measure of success is the overall percentage of *all* students achieving Level 3 or 4 — the students who are considered to have mastered the curriculum — *and* if all or virtually all of the

Table 5.1: ***The Exemption Process, academic years 1998–99 to 2001–02***

	1998–99	1999–2000	2000–01	2001–02	Null Hypothesis: Equality of Exemption Rates across Years
	(% average exemption rate)				*(probablity value)*
Grade 3					
Reading	6.56	6.90	5.97	5.61	0.00
Writing	5.33	5.99	5.32	5.07	0.00
Mathematics	4.84	5.36	4.68	4.43	0.00
Null hypothesis: Equality of exemption rates across tests *(probability value)*	0.00	0.00	0.00	0.00	
Grade 6					
Reading	4.36	5.12	3.92	4.22	0.00
Writing	4.15	4.98	3.86	4.12	0.00
Mathematics	4.18	4.99	3.84	4.01	0.00
Null hypothesis: Equality of exemption rates across tests *(probability value)*	0.46	0.71	0.90	0.47	

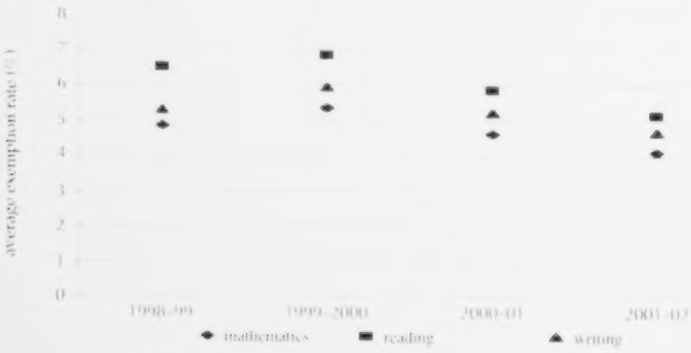
Note: Shading indicates that the null hypothesis cannot be rejected at a 5 percent level of significance.

exempt students would not have achieved those levels, then variation in the percentage of exempted students is not a critical issue in consideration of results for the province as a whole. In any case, all 24 exemption rates are small, and the overall exemption rate is not large enough to have an impact at the provincial level.

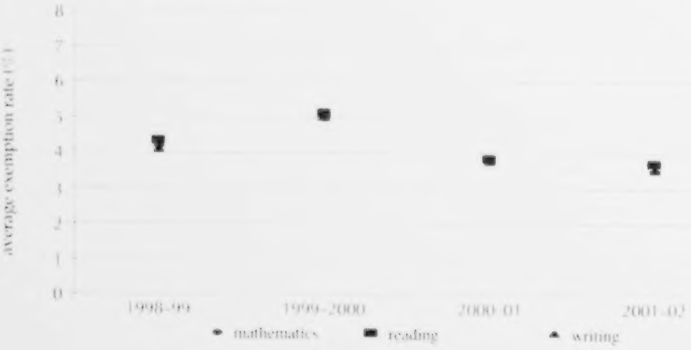
Most of our interest in the exemption process is, in fact, indirect. The exemption data provide small clues about the assessment process itself. The variation in the exemption rate over time, even in a very large sample, is in some ways completely understandable. It merely signals the difficulty of designing and applying a perfectly consistent exemption process to an assessment. The variation in exemptions leads to the conclusion that any comparison of results between schools requires using Method 1, rather than Method 2. This is the exact exercise I undertake in Chapter 6. First, however,

Figure 5.1: ***Average Exemption Rates by Assessment and Year, academic years 1998–99 to 2001–02***

Panel A: Grade 3 assessment



Panel B: Grade 6 assessment



it is worth looking at the relationships between the socio-economic characteristics that describe a school community and the exemption rate at a school, since such an exercise allows one to measure variations in exemption rates across school boards.

Do School Boards Have Different Exemption Rates?

The methodology I used in Chapter 4 to consider variations in achievement results can also be used to measure the association between exemption rates and the socio-economic variables that describe a school community, as well as whether schools with the same socio-economic characteristics in different school boards have the same exemption rate. Indeed, it would not be surprising if such board effects existed. All boards get the same exemption instructions from EQAO, but principals are then left to interpret these instructions. A "group" of principals at the same board, who know each other and likely received some common training on the administration of the assessments, would very likely make similar decisions. This analysis is designed to measure the strength of such effects and finds that, in fact, there is substantial variation in exemption rates at similar schools in different boards. This again suggests that Method 2 results are not comparable across boards.

A simplified version of the equation estimated is:

$$EX_S = \alpha + \hat{\delta}_{board} D_{board} + \beta_X{}^1 X^1{}_S + \beta_X{}^2 X^2{}_S + u_S . \qquad (5.1)$$

The sample consists of 3,300 schools. D_{board} is a dummy variable that takes on a value of 1 for an observation from a specific board, and zero in the observation at a school in any other board. $X^1{}_S$ and $X^2{}_S$ represent the values at school S of the same socio-economic indicators used in Chapter 4. Exemption rates are so similar at a given school in each grade in a given year and at the same school across the four years that it is necessary to average across these values over the four years and create one average exemption rate per grade per school. As indicated in the technical appendix to this chapter, it is not possible to estimate equation (5.1) for all schools and boards. To estimate an effect of membership in a board on the exemption rate (the value of $\hat{\delta}_{board}$), there must be enough schools in a board to estimate the effect of belonging to the group. Some boards, however, have fewer than ten schools and are excluded from the estimation of equation (5.1).

Table 5.2 presents the results from the estimation of equation (5.1). The list of socio-economic variables that describe school communities used in Chapter 4 forms the list of right-hand variables in the table, the $X^1{}_S$ and $X^2{}_S$ variables used to estimate equation (5.1). Unlike

Table 5.2: *Socio-Economic Determinants of Exemption Rates at Schools and Testing for Board Effects on Exemption Rates*

Variable	Grade 3 Coefficients (Standard Error)[a] Without Board Dummies	Grade 3 Coefficients (Standard Error)[a] With Board Dummies[b]	Grade 6 Coefficients (Standard Error) Without Board Dummies	Grade 6 Coefficients (Standard Error) With Board Dummies
Single-parent families (%)	0.007 (0.01)	-0.01 (0.01)	0.02 (0.02)	-0.0002 (0.01)
Official language spoken at home (%)	-0.08 (0.01)	-0.05 (0.01)	-0.05 (0.02)	-0.01 (0.01)
Aboriginal descent (%)	0.01 (0.04)	0.1 (0.05)	0.03 (0.04)	0.08 (0.06)
Recent immigrants (in Canada < 5 years) (%)	-0.1 (0.03)	-0.08 (0.06)	0.00003 (0.03)	-0.01 (0.04)
Living in detached home (%)	-0.02 (0.006)	-0.03 (0.006)	-0.008 (0.006)	-0.01 (0.01)
Log of income	-2.85 (0.54)	-2.92 (0.65)	-1.79 (0.55)	-2.45 (0.69)
Moved in past year (%)	0.14 (0.04)	0.19 (0.04)	0.05 (0.04)	0.11 (0.04)
Moved in past 5 years (%)	-0.01 (0.01)	-0.04 (0.01)	0.02 (0.01)	-0.01 (0.01)
Unemployment rate (%)	0.14 (0.04)	0.1 (0.04)	0.08 (0.04)	0.09 (0.04)
Adults without high school diploma (%)	0.03 (0.01)	0.07 (0.02)	0.004 (0.01)	0.01 (0.01)
Adults with some university (%)	0.05 (0.02)	0.05 (0.02)	-0.01 (0.02)	-0.02 (0.02)
Total enrollment	0.000 (0.00)	-0.001 (0.0005)	-0.001 (0.0004)	-0.001 (0.0004)
R^2	0.18	0.27	0.14	0.24
Number of schools	3,352	3,352	3,352	3,352
Number of board dummies significantly different from zero[c]	OLS Tobit Log	21/59 24/59 21/59	OLS Tobit Log	23/59 24/59 14/59

[a] Shaded cells contain coefficient estimates that are not statistically different from zero at the 95 percent confidence level.

[b] A sufficiently large number of schools in a board is needed to estimate the value of δ_{board}; accordingly, boards with fewer than ten schools are excluded from the estimation of equation (5.1).

[c] There are 59 coefficients estimated on board dummy variables. The first value counts the number of dummy variables significantly different from zero — that is, whose exemption behaviour is significantly different from that of the base board, the Waterloo Region District School Board.

Chapter 4's results concerning the association between assessment results and socio-economic variables, however, many of the coefficients that measure the association between the exemption rate and socio-economic variables do not differ statistically from zero. Single parenthood does not affect exemptions in either grade. Exemptions fall when more people in the school community speak an official language at home, but the effect is not statistically significant in either grade. Exemptions rise in Grade 3, but not in Grade 6, when there are more aboriginal students. There are actually fewer exemptions when there are more recent immigrants in Grade 3, while there is no association between recent immigrants and exemptions in Grade 6. There are fewer exemptions when there are more detached homes, but this effect is not statistically significant in Grade 6. Exemptions rise in both grades when more households in the school community have moved in the past year, but not when more households in the community have moved in the past five years. Finally, exemptions rise in both grades when the unemployment rate is higher, and fall in both grades with an increase in (the logarithm of) income.

None of these results is surprising. What is surprising, however, are the coefficients on the parental education variables describing the school community. In Grade 3, if there are more adults without a high school diploma, exemption rates rise, but if there are more adults in the school community with some university education, then exemption rates rise even more. Neither effect is significant in Grade 6. These results suggest two puzzles: why would the community educational level matter in Grade 3 but not in Grade 6? and why do better-educated parents of students in Grade 3 seem to request an exemption more often?

As a final note, smaller schools have slightly lower exemption rates in both Grade 3 and Grade 6, but the effect is very small. The R^2 in the ordinary least squares estimates of equation (5.1) are, without board dummy variables, 0.18 in Grade 3 and 0.14 in Grade 6. These are relatively small values of R^2. The technical appendix discusses the need to estimate equation (5.1) using other statistical techniques beyond ordinary least squares. These different techniques do not change the results.

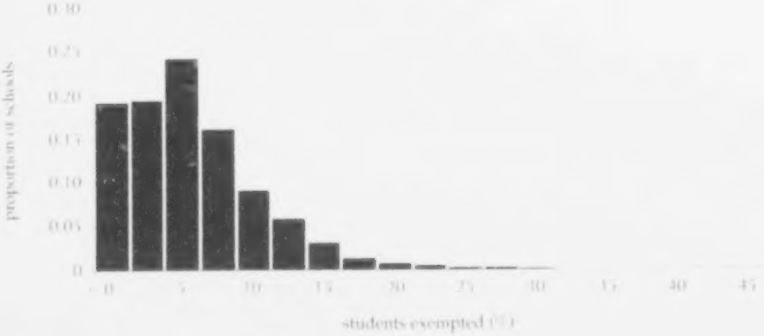
The remaining issue is to measure whether the exemption process is consistent across different boards. Board effects are measured by estimating the values of the coefficients δ_{board}, one coefficient value per board. These are the coefficients on the dummy variables that take on a value of 1 if school *S* comes from a specific board and zero otherwise. Schools at 60 different boards are included, leaving 59 values of δ_{board} to be estimated.

Table 5.2 reports the number of values of value of δ_{board} that are statistically different from zero in different estimates of equation (5.1). As described in the technical appendix, three estimators are used. A significant value of δ_{board} means that, in schools of that board whose socio-economic variables describing the school community have the same value, exemption rates differ systematically from those of schools in the Waterloo Region District School Board — a board chosen because it is of medium size with a wide variety of schools, both rural and urban, rich and poor.

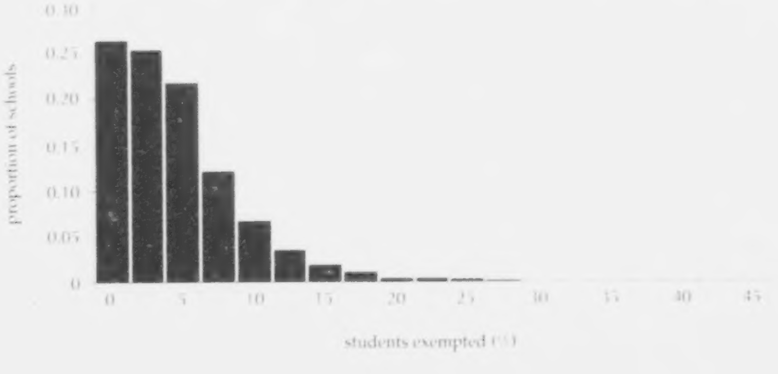
The table presents substantial evidence in favour of board-level variations in the application of exemption criteria. As can be seen in the last three rows, depending on the exact estimation procedure, about one-third of the 60 boards behave *differently* from the Waterloo board. This is not to say that the exemption behaviour of those boards or of the Waterloo board is either good or bad, only that exemption behaviour differs across boards at schools with the same socio-economic characteristics. At the same time, these differences, although of statistical significance, are not large, with δ_{board} ranging only from 2.5 to 11.5 in absolute value terms. This means that exemption rates at an average school in one of these boards might be as much as 11 percentage points higher or lower than that of an average school in the Waterloo board. I stress that these very high numbers apply to only a few boards. The variation in exemption rates across schools, shown in Figure 5.2, is actually much larger, as would be expected. Nevertheless, whatever the reasons schools exempt students from the assessments, they are not very strongly correlated with socio-economic factors that describe

Figure 5.2: ***Distribution of Student Exemptions from Assessments***

Panel A: Student exemptions from the Grade 3 assessment



Panel B: Student exemptions from the Grade 6 assessment



the school community nor with the school's being part of a specific board. In short, the variations in exemptions are not readily explained with a simple set of variables.

Conclusions

In this chapter, I have offered a short analysis of the rates at which Ontario elementary school students are exempted from taking the standardized assessments. There is clear evidence that exemption rates vary over time. There is also some evidence that exemption rates vary across boards at schools with similar socio-economic community characteristics; however, the variations in exemption rates are not closely associated with variations in socio-economic variables.

These results are not surprising. None of the analysis in either Chapter 4 or in Chapter 6, which follows, depends heavily on the consistent application of exemption criteria as long as we believe that most exempted students would not have performed at Level 3 or 4 in any case.

The important policy conclusion from this chapter is that attention should be directed away from presenting assessment results according to EQAO's Method 2, which leaves results open to manipulation if the exemption criteria are applied inconsistently. Indeed, there is clear evidence of inconsistency from one academic year to the next and quite strong evidence of some inconsistency across boards. If a school or a board set as a goal a larger percentage of examinations *actually written* to score at Level 3 or 4, then the logical strategy would be to exempt more students who would not achieve those levels. This would be a simple and highly effective strategy to raise results in an environment in which the Method 2 calculation of results is emphasized. Although there is no direct evidence that such a strategy is in place, one would likely emerge rapidly if Method 2 results were to become a public policy focus. The obvious and clear solution is to refrain from presenting or emphasizing Method 2 results and to restrict the analysis to the percentage of students scoring at Level 3 or 4 when taken as a percentage of all students. This task continues in the next chapter.

Chapter Summary

Some Ontario elementary school students are exempted from the Grade 3 and Grade 6 assessments undertaken by the Education Quality and Accountability Office (EQAO). Students can be exempted when, even with all possible accommodations, "the student would be unable to participate productively and/or where the student's participation would be harmful."

The overall exemption rate is quite low — falling between 4 and 7 percent — but it varies across years and across assessments. There is also some evidence that exemption rates vary across boards — that is, two schools with the same socio-economic characteristics but located in different boards can have systematically different exemption rates.

Exemption rates are important because some users of school assessment results emphasize EQAO's "Method 2" presentation of results, which are open to manipulation. In this methodology, the key variable is not the percentage of *all students* at a school that achieves at Level 3 or Level 4 but the percentage that *actually wrote the assessment* at a school that achieves at those levels. Thus, a school or board that wanted to improve its apparent assessment results using Method 2 could do so simply by exempting more students who would not score at Level 3 or Level 4. In the logical extreme, "perfect" Method 2 results could be obtained by exempting all students except those who would score at Level 3 or Level 4. Most school boards, EQAO itself, and most users of EQAO data do not make use of Method 2 results. However, the Fraser Institute uses Method 2 in its annual report card, which I argue is an inappropriate way to present assessment results.

Technical Appendix

The estimation of equation (5.1) contains several new statistical issues. Figure 5.2 shows that the distribution of exemptions is very highly skewed to the right — that is, a few schools have very high exemption rates — while many (131) schools in the sample have an exemption rate of zero, even when averaged over the four academic years. Both characteristics of exemption rates present particular problems in the analysis when using ordinary least squares. To be consistent with the analysis in Chapter 4 and Chapter 6, Table 5.2 presents ordinary least squares estimates of equation (5.1). Chapter 4's list of socio-economic variables that describe the school community forms the list of right-hand variables to include in the equation. Notice, first, that the R^2 is quite low, particularly for the equations without board dummy variables. Many of the coefficients measuring the association between the exemption rate and socio-economic variables (the shaded areas in the table) are not statistically different from zero. There is, however, quite a sharp contrast between the amount of shading in Table 5.2 and that in Tables A4.1 and A4.2. Variations in exemption rates are not easy to explain.

Because of the distribution of exemption rates, ordinary least squares may not be the most desirable estimator to use to discover the association among community variables, board dummy variables, and exemption rates. Accordingly, I re-estimated the equations both by using a Tobit model — which accounts for the truncation of observations at zero — and by the transformation of the dependent variable by taking its natural logarithm.[4] Since the schools of 60 boards are included, there are 59 estimates of δ_{board} using three specifications of equation (5.1). The last row of Table 5.2 reports the number of board dummies that are statistically different from zero in each specification of equation (5.1). A significant coefficient on a board dummy variable means that, in schools in that board with the same values for the socio-economic variables that describe the school community, exemption rates differ systematically from those in schools in the base case board, the Waterloo Region District School Board. The coefficients in Table 5.2 provide substantial evidence in favour of board-level variation in the application of exemption criteria. Depending on the exact model, about one-third of the boards behave differently from the Waterloo board. When board dummies are added to the estimated equation, the R^2 increases substantially. This is further evidence of some inconsistency in the application of the exemption process across boards.

---

[4] I have not included coefficient estimates for these versions, since they do not differ very much from the coefficients reported. The transformation of exemption rates by taking logarithms drops from the data schools with exemption rates of exactly zero.

6 Ranking Schools Using the EQAO Assessment Data

The results of the assessments of Ontario's elementary school students undertaken by the Education Quality and Accountability Office (EQAO) are put to a variety of uses. They provide curriculum guidelines for teachers and students. They fill gaps in the curriculum by creating explicit expectations. And, if one is willing to assume consistency in design and grading over time, they provide an index of overall achievement by students and at specific schools and boards.

The assessments are intended to be an absolute measure of the level of performance by students against a well-defined standard. Not all participants in the process would agree, however, that this goal is being achieved. Yet, even if the assessment process is unable to ascertain clearly if Ontario students have improved over time, the results provide feedback on strengths and weaknesses of both students and schools in specific areas. This feedback comes from a source one step removed from the student's individual teachers, as other teachers who grade the assessments both provide and gain insight into the abilities of a very large sample of students.

All of these aspects of the assessment process are useful and relate to the goals of the EQAO process in the Ontario elementary school system. Now, in this chapter, I explore the use of the assessment results to rank schools — that is, to help decide if teachers and staff at one school are doing a "better" job than teachers and staff at another school.

Using the data made available by EQAO, Ontario's elementary schools can be ranked according to two methods.[1] One method, called an "absolute" ranking, simply ranks schools according to the percentage of their students performing at Level 3 or Level 4 on the assessments. As we have already seen, schools with strong assessment results tend to be those that draw their students from communities where parents are better educated, more affluent, and less mobile, where aboriginals are a small proportion of the population, and where a large proportion of students live in two-parent families in detached homes. The well-known rankings used by the Fraser Institute[2] are based on schools' absolute assessment results. This methodology, in my view, is not very useful in measuring the effectiveness of either the educational process or the teaching staff at those schools.

---

[1] There is, in fact, a third method of assessing the effectiveness of schools, known as the "value-added" approach, but it cannot be implemented using EQAO data in their present form. For a description of the value-added approach, see Thomas (1998), particularly pp. 101–03, which describes a value-added pilot project in the English county of Lancashire. For a more recent survey, see Ladd and Walsh (2002).

In the United Kingdom, national individual student numbers were created in 1999 to facilitate the handling of individual student data. The key to a value-added assessment of schools is that assessment data must be available at the level of many individual students at two different points in time. Suppose assessment results were available for both Grade 3 and Grade 6 for many students. One could then ask, given the socio-economic background of either specific students or all students in a school, at which schools (and even, if the data were available, for which individual teachers) were learning gains the largest. Such data would be ideal to use to compare the effectiveness of schools and teachers. However, although partial or complete data sets of this type exist in the United Kingdom, in some U.S. states (North Carolina, South Carolina, Tennessee, and Texas), and in British Columbia, they are not currently available in Ontario.

[2] See Cowley and Easton (2003); a 2004 edition using the same methodology is now available.

A more useful and relevant methodology, however, is a "relative" ranking system that identifies schools that perform better than *similar schools* on the same assessments — in other words, a ranking that compares schools that serve communities with similar socio-economic characteristics.[3] Accordingly, in this chapter, I present details of a method that creates a relative ranking of schools and that clearly identifies schools that are outstanding relative to other, comparable schools. This relative ranking method takes into account the socio-economic context of the school community and, importantly, adjusts assessment results for that context.

The chapter closes with a comparison of the schools in the Waterloo Region District School Board using both the Fraser Institute's absolute rankings and the relative rankings produced in this study. The reader can then decide which ranking methodology is more informative and useful. Rankings of all Ontario elementary schools for which sufficient data exist can be found on the C.D. Howe Institute's web site at: www.cdhowe.org.

A Reminder of the Variables Used in the Analysis

In Chapter 4, the focus was on the percentage of *all* students at a school that achieved at either Level 3 or Level 4. This, in EQAO's terminology, is the Method 1 presentation of assessment results, and is a variable that could be viewed as a school's "success rate," to the extent that the choice of the benchmark of Level 3 or Level 4 indicates success — although it should again be noted that not all participants in the process agree with that choice of language. Chapter 4's more neutral label for this variable was "Category One."

---

[3] I should make it clear, however, that absolute scores may be very useful for some purposes. *If the assessment instrument remains comparable*, an overall increase in provincial scores from, say, 50 percent of students at Level 3 or better to 75 percent of students at Level 3 or better on a comparable test would be wonderful news. But the comparison of absolute scores between schools as a measure of school "success" is not, in my view, the most useful comparison with which to measure educational effectiveness in most circumstances. Even if the provincial average score increased, one would still want to measure a school's relative success within an overall increase in scores.

In this chapter, to use even more simple language that does not contain an element of judgment on the issue, I refer to this variable as an "assessment score."

As we saw in Chapter 4, enormous variations exist across schools in assessment scores, across the different types of assessments, across years on the same type of assessment, and across schools in the same year. In looking for schools that outperform other, similar schools, we need to identify patterns across these assessment scores relative to two elements of the context in which schools operate — namely, the year in which the assessment was undertaken and the type of test that was administered. In other words, to make a "fair" comparison of outcomes across schools, it is necessary to compare scores on the same assessment from the same year, otherwise we may end up comparing a school writing in a year with a "hard assessment" with one writing in a year with an "easy assessment." Moreover, since success rates on the mathematics assessment differ from those on the other assessments even within the same year, this aspect of context must also be included in the analysis.

In Chapter 4, we also saw that various socio-economic variables that describe the school community are strongly associated with school assessment scores. To review, those socio-economic variables are as follows:

- the percentage of families that is headed by a single parent;
- the percentage of households that speaks an official language at home;
- the percentage of the school community that identifies as aboriginal in origin;
- the percentage of the school community resident in Canada for less than five years;
- the percentage of the school community born outside Canada, the United States, or the United Kingdom;
- the percentage of the school community living in detached homes;
- the logarithm of average household income in the school community;
- the percentage of households that moved in the past year;

- the percentage of households that moved in the past five years;
- the unemployment rate;
- the percentage of adults in the school community without a high school diploma;
- the percentage of adults in the school community with some university education; and
- total enrollment in the school.

Even though it is rarely of statistical or practical interest, enrollment is on the list because of the educational community's interest in the issue of school size.

Creating a Measure of the School's Relative Performance

The process of developing a measure with which to compare the performance of schools in the context of both variations in average assessment results and the socio-economic characteristics of the school community is the work of this section.

Creating a measure of relative rankings requires three steps. The first step is to take into account variations in average assessment results. The second step is to take into account variations in results that can be associated with variations in the socio-economic context variables that describe a school community. The third step is to "normalize" the measure of a school's relative ranking to provide a more straightforward and unit-free interpretation of that number.

The First Step: Accounting for Variations in Average Assessment Results

In Chapter 4, the unit of observation was an assessment score in reading, writing, or mathematics in either Grade 3 or Grade 6 at a particular school in one of four academic years, 1998–99, 1999–2000, 2000–01, or 2001–02. In this chapter, I create the unit of observation using all four years of results from a school for either Grade 3 or Grade 6.

In Chapter 4, we saw that it was useful to analyze Grade 3 and Grade 6 assessment scores separately because the association of higher scores with specific socio-economic community characteristics is slightly different from one grade to the other and because these scores have different average values. In this chapter, because the unit of observation is a grade at a school, the scores on Grade 3 (or Grade 6) assessments at a single school — up to 12 scores in the 12 different assessments over the four years — must now be transformed into a single number. We also saw in Chapter 4 that the provincewide average level of school scores varies over the years and across assessments for reasons that are not straightforward to determine — that is, for reasons that are essentially unknown. For our purposes, it is not important that these reasons are unknown; however, in order to make a valid comparison of schools, it is important to remove the variations in average assessment scores across the province.

The first step is to calculate the average score on each assessment in each year across all schools. The next step is to subtract the average value on a particular assessment for all schools in the province from the score that each school obtained on that assessment. If this variable has a negative value, the school had a lower percentage of students scoring at Level 3 or Level 4 on that assessment in that year than the average for all schools in the province. A positive value means the opposite. The unit of the variable is the percentage of *all* students at a school. A value of 10 means that the school had 10 percentage points more students at Level 3 or Level 4 (as a percentage of all its students) than did the average school on that particular assessment. The average value of this variable over all schools on a specific assessment is zero by construction. Then, for each school, I averaged the values of this variable on (up to) 12 assessments as the measure of school performance in either Grade 3 or Grade 6.[4] This variable, called the "adjusted school score" represents the school's achievement level fully adjusted for variations over time in the difficulty of the assessment, in the way students were graded, or in achievement levels for the province as a whole. Using this scaling, if the level of achievement for the province as a whole increased, the variable is higher only if a school's assessment scores increased at a faster rate than that of the province as a whole. The "adjusted school score" at school S is denoted S_S. Each school usually has two adjusted scores, one for Grade 3 and one for Grade 6 (some schools have one grade but not the other). This methodology removes from the school-level assessment results any effects of the type of assessment written — whether reading, writing, or mathematics — as well as effects associated with the year in which the assessment was written.

---

[4] The mean of this variable across all schools for a given grade is not quite zero, since some schools had fewer than 12 assessments and data are missing for some schools for some years. In addition, if a school has fewer than 15 students...

The Second Step: Accounting for Variations in Socio-Economic Context

The next step in comparing and ranking schools is to take into account variations in the characteristics of the community in which each school operates. We saw in Chapter 4 an extensive investigation into the association between socio-economic variables that describe the school community and assessment results at schools. This chapter uses similar techniques to measure the association between those same socio-economic variables and the adjusted school score variable described above.

As in Chapter 4, it is quite difficult to present the methodology of this analysis without the use of equations. Accordingly, the reader who is uncomfortable with this form of presentation may wish to skip ahead two sections to the ranking of schools using the example of the Waterloo Region District School Board.

Note 4 - cont'd.

...in a given year, the data may be suppressed for privacy reasons. Moreover, in some years, French immersion schools were given the choice of opting out of some assessments.

The equation,

$$S_S = b_0 + b_1 X^1{}_S + b_2 X^2{}_S + b_3 X^3{}_S + b_4 X^4{}_S + u_S, \qquad (6.1)$$

describes adjusted scores at school S, where the values of the coefficients b_1, b_2, b_3, and b_4 represent the average association between the adjusted scores at all schools and the value of X^1, X^2, X^3, and X^4, respectively. (The estimates of these coefficients are reported in Table A6.1 in the technical appendix.) An equation is estimated using the adjusted school scores from the Grade 3 assessments and another equation is estimated using the adjusted school scores from the Grade 6 assessments. These equations are estimated for all schools for which at least six assessment results are available to form the average adjusted result — a total of 3,583 schools containing Grade 3 and 3,402 schools containing Grade 6.

The coefficients on the X variables represent the association of an X variable with an adjusted school score. The signs indicate the direction of the relationship. When $X^1{}_S$ changes by one unit, then S_S is predicted to change by b_1 units. As in Chapter 4, it is interesting to compare the coefficients on the same variables in both the Grade 3 and Grade 6 equation, and to look at the sign and statistical significance of the estimates of these coefficients.

In estimating these equations, we find the following:

- Scores on the assessments for both grades decline as the percentage of single-parent families in the school community increases, and the effect is significantly larger in Grade 6 than in Grade 3.
- Scores decline as the percentage of the community that speaks one of the official languages at home increases, although the effect is not statistically significant in the Grade 6 equation.[5]

---

[5] This result was obtained in Chapter 4 using both the census and EQAO-based measures of language at home. There must be an interaction between language and immigration in the data so that the positive effects of immigration spread into the language measure.

- Scores decline as the percentage of aboriginals in the school community increases, with the effect being about the same for both the Grade 3 and Grade 6 assessments.
- Scores in both grades increase as the percentage of recent immigrants increases, although the association is weaker and not statistically significant in Grade 6.
- Higher incomes are strongly associated with higher adjusted scores in both grades by approximately the same amount.
- Scores increase as the percentage of the school community that lives in detached homes increases, with the association being about the same for both grades.
- Scores decline as unemployment in the school community increases, again with the association being about the same for both grades.
- Scores decline slightly on the Grade 6 assessments as school size increases, although the effect is not of much practical interest. There is no measurable effect on the Grade 3 adjusted scores. If a school were to increase in size by 500 students (a very large increase), the percentage of students performing at Level 3 or 4 in Grade 6 is predicted to fall by 0.35 of a percentage point.

As in Chapter 4, the most interesting coefficients are those on the two variables that describe the educational context of the school community. Adjusted scores on the Grade 3 assessments decline as the percentage of adults without a high school diploma increases, but the Grade 6 assessments remain unaffected. Scores increase significantly for both grades — but particularly for Grade 6 — as the percentage of adults in the school community with some university education increases.

In view of the fact that these socio-economic variables appear to have different effects on the Grade 3 assessments than on the Grade 6 confirms the importance of estimating equation (6.1) separately for the two grades.

By now, the reader will likely have realized that this exercise yields results that are very similar to those obtained in Chapter 4 using observations of class-level assessment results. Why, then, is it useful to re-estimate these equations using adjusted scores across a sample of schools? Part of the reason for doing so is the revelation that the explanatory power, the R^2, of both the Grade 3 and Grade 6 equations (presented in Table A6.1 in the technical appendix) is quite high — indeed, it is much higher than in similar equations estimated in Chapter 4. This is an important result. Recall that the R^2 measures the amount of variation in the adjusted score that is associated with variations in the school's socio-economic context. The value of the R^2 must lie between 0 (there is no association of socio-economic context variables with adjusted scores) and 1.0 (all variations in adjusted school scores are associated with variations in socio-economic context variables). The estimated equations in this chapter reveal that the R^2 for the Grade 3 adjusted scores is 0.40 and for Grade 6 it is 0.44. These R^2 values are substantially higher than the values of 0.25 and 0.28, respectively, that were obtained in Chapter 4 using observations of grade-level assessment results.

Why is this increase in the R^2 in the two sets of equations so important for understanding the EQAO assessment results? Recall that, to create the left-hand variable in equation (6.1), the adjusted school scores are averaged over many assessment results at each school over a four-year period. The high R^2 values of 0.40 (for Grade 3) and 0.44 (for Grade 6) mean that 40 or 44 percent of the variation in the school's average score over many observations within the four-year period is associated with variations in the variables that describe the school's socio-economic context. In other words, these context variables have more explanatory power when assessment results are averaged over a number of years than they do over a single year of assessment results.

This result is both sensible and entirely consistent with those of other studies. Some of the literature on school testing emphasizes the small sample size that a class (or several classes) constitutes in a given year (see, for example, Kane and Staiger 2001; and Chay, McEwan, and Urquiola 2003), and as the next chapter makes clear, teachers and principals repeatedly make exactly the same point. Students in a given year on a given assessment at a specific school are often a fairly small group — perhaps 20 to 30 children in a single class or 30 to 40 in two classes or a split grade — and the characteristics of such a small group are highly variable from year to year. This variability introduces substantial fluctuations from year to year in a school's adjusted score. Teachers often have a good idea what a given class's EQAO assessment results are likely to be, which they express in statements such as "this year's class is weaker (or stronger) than last year's." Thus, a "small" fluctuation in the characteristics of a single class often leads to a large fluctuation in the school assessment score on a particular assessment from year to year.

Consider a practical example. Suppose that, in a class of 20 students, two students were to "drop" from Level 3 to Level 2 from one academic year to the next; the school assessment score, or the adjusted school score, would then decline by 10 percentage points. This outcome might be unrelated to any activity undertaken or choices made by either the class or the school. It would just happen. This underscores the importance of averaging the dependent variable S_S in equation (6.1) over the four years and over all 12 assessments (four each in reading, writing, and mathematics), which then removes some of the random fluctuations in the academic quality of students over the four years.[6] Thus, this chapter compares schools on the basis of the performance, not of a single class, but of many classes in all three subjects, and not in just one year but over a four-year period, otherwise the sample would be just too small.[7] This approach has very practical implications for the interpretation and presentation of assessment results, an issue to which I return in Chapter 8.

---

[6] Note, however, that averaging over the four years at a school might or might not remove random fluctuations in the quality of teachers since the classes might not have had the same teacher over that time. No such data are currently publicly available, although schools boards are likely to have such information. Clearly, there is scope for further statistical work in this regard.

[7] It is essential in future work to determine the number of years of results that should be averaged to create a measure of school performance. The use of too few years increases the amount of variation due to random changes in class ability. The use of too many years (say, ten) increases the likelihood that the factors that contribute to performance at a particular school will change too much over time for their identification to be of practical value.

Finally, it is important that the R^2 in the estimates of equation (6.1) is not 1.0. In other words, although the school community's socio-economic context is important in explaining variations in the school's adjusted assessment results, it does not explain all of those variations. This finding is crucial in that it sharply contradicts the assertion of some participants in the assessment process that only schools that draw students from communities with strong socio-economic characteristics can score well on the tests.

The Third Step: Measuring School Performance after Accounting for Variations in Socio-Economic Context

The third and final step in creating a relative ranking of schools is the most complex. In it, we use the estimates of equation (6.1) to create a numerical measure with which to compare the performance of schools after accounting for the socio-economic context in which they function. For the technical details of how this measure is created, I refer the interested reader to the technical appendix at the end of this chapter. Suffice it to say here that one needs to isolate that part of the school's adjusted assessment score that is not related to variations in the socio-economic characteristics of the school community. Then, the school's score can be compared with the average for all schools in the province.

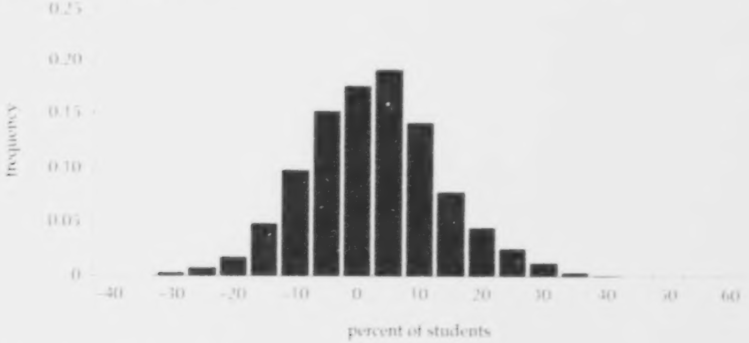
For each grade at each school, a value, u_S, is calculated from equation (6.1) as the difference between the school's actual adjusted school score and its predicted adjusted school score. If that value is zero, the school's performance in that grade is exactly as predicted by the combination of variables that describe its socio-economic context. In a sense, such a school is behaving as the "average" school in the province. If the value of u_S is positive, the school is performing better than other schools after taking into account variations in its socio-economic context and the provincewide variation in the average results on different assessments written at that school. To repeat, the value of u_S clearly represents variations in school performance after socio-economic variables have been taken into account. If two schools have the same predicted adjusted score, but one school's u_S has a positive value while the other's u_S has a value of zero, then the school with the positive value is a "better" school.

One final step is needed to help interpret these values of u_S. The units of u_S in equation (6.1) are percentages of students at a school — that is, a value of 10 indicates that, in this school, 10 percentage points more students achieve at Level 3 or Level 4 than were predicted to do so at an average school in the province. In a class of 20 students, 10 percentage points represents two students. But because this value is defined over four years, we now have a total of 80 students, with 8 more students performing at Level 3 or 4 than in a group of 80 students at an average school in the province. Thus, a value of $u_S = 10$ is a substantial variation in outcomes. Having established that a value of 10 is large enough to be of interest in an educational sense, the final step in the construction of relative school rankings is to ask whether it is common or uncommon to observe values of u_S that approximately equal 10.

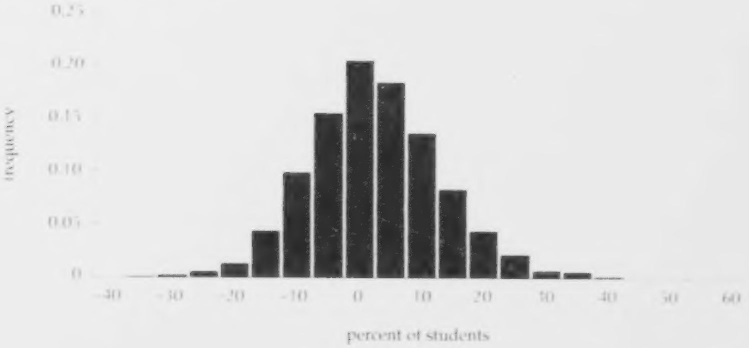
The most common statistical method used to compare variables is to take values of a variable and create a "normalized variable." Figure 6.1 presents pictures of the values of u_S, from the Grade 3 and Grade 6 equations, respectively. Visually, the two panels of Figure 6.1 look like "normal" distributions: they are bell-shaped distributions, where most of the observations are close to zero. This means that it is most likely that observations will occur close to zero and least likely that observations will occur far away from zero. About half the values are negative and the other half positive. Figure 6.1 illustrates that, at the typical school, one is quite likely to observe a value of u_S of +5 or –5 percentage points. A value of u_S of +20 or –20 percentage points, however, would clearly be very unusual. It would mean that 20 percentage points more students on average in every year (4 students per year in a class of 20, or 16 students over four years) do much better at this school than a similar sized group at other schools, including those with similar socio-economic

Figure 6.1: ***Difference between Actual and Predicted Adjusted Assessment Scores in Ontario Elementary Schools***

Panel A: Grade 3 students



Panel A: Grade 6 students



settings. This is a very large difference. But how likely is it that one will observe any particular value of u_S?

The way to determine the likelihood of any particular value is to calculate the standardized value of u_S. Think of rolling a pair of dice. It is very common to roll a 7, and is analogous to observing a value of u_S of +5 or –5. On the other hand, it is much less common to roll a 2 or a 12, which is analogous to observing a value of u_S of +20 or –20. While many of us may be familiar with the statistical properties of dice rolls, the statistical properties of u_S are much less familiar. Standardization is a way to create familiarity.

Here, we create and use a variable called the Z-score. For most schools, there are two Z-scores, one for Grade 3 and the other for Grade 6. Using tables of the standard normal distribution in any statistics book, one can translate Z-scores into percentiles that tell us how unlikely is a given observation of u_S. Thus, if we observe a school with a Z-score of +1.29, that means its adjusted assessment score is higher than that of 90 percent of schools (after taking into account variations in their socio-economic settings) — that is, 90 percent of the observations of Z on a normal distribution are less than 1.29. If we see a school with a Z-score of +2.33, that means its assessment results are better than those of 99 percent of schools; similarly, a Z-score of –2.33 means that the school is *worse* than 99 percent of schools.

In short, Z-scores are a convenient way to create relative rankings of schools now fully adjusted for both the socio-economic context of the school community and the year and type of assessment. The Z-score allows us to place the school on a percentile ranking of all schools in a convenient and understandable way. A high and positive Z-score indicates a better school; a low and negative Z-score indicates a school with lower results than comparable schools. The ability to associate a percentile ranking and the normal distribution with a Z-score is useful in the sense that, although the great majority of schools are likely to be close to one side or the other of zero, very high or very low Z-scores are unlikely to be chance events — indeed, schools that are so much better or worse than other schools will be of considerable statistical and educational interest.

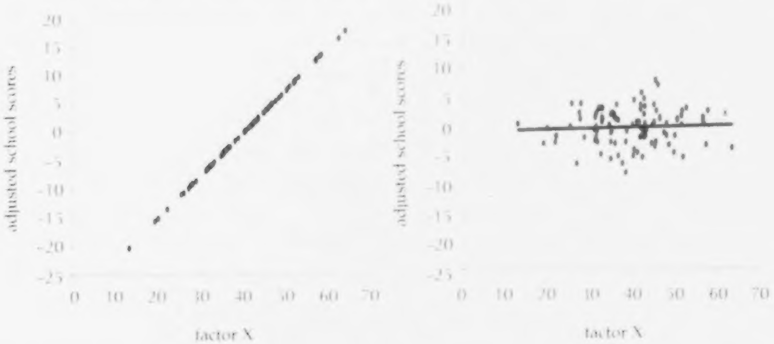
To Summmarize:
A Graphical Example of School Rankings

The relative ranking process is illustrated in the four panels of Figure 6.2. For the sake of simplicity, this example uses only one numerically measured socio-economic factor, called factor X, that might be associated with adjusted school scores. Panels A, B, and C present three possible scenarios for the relationship between X and the adjusted school score. Each dot represents an observation of the

Figure 6.2: *A Visual Explanation of Relative School Rankings*

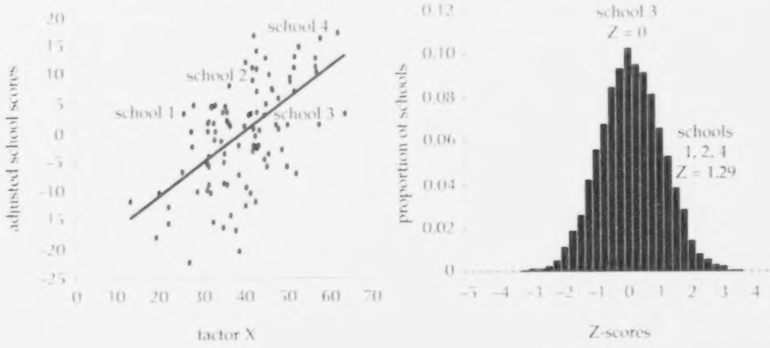
Panel A: Factor X captures all the variation in adjusted school scores

Panel B: No relationship between factor X and variation in adjusted school scores



Panel C: Positive relationship between factor X and variation in adjusted school scores

Panel D: Placing Z-scores on a normal distribution



adjusted school score and the value of factor X at a school. The predicted values of factor X for each school are represented by the lines drawn on panels B and C (a line on panel A would exactly cover the plotted observations), while the estimated values of u_S are represented by the vertical distance between the line and the dot for each observation at a school.

Panel A represents the situation where factor X captures *all* of the variation in adjusted school scores. In other words, the R^2 is 1.0, and all values of u_S are zero, meaning that adjusted school scores are perfectly associated with variations in factor X. Thus, for example, if X were family income, then only schools that draw their students from communities with high income levels would score well on assessments. Note that, in panel A, all values of u_S are zero.

Panel B represents a scenario where there is no relationship between factor X and the adjusted scores of schools, so that knowing the value of factor X for any particular school tells us nothing about what the school's adjusted score is likely to be. In this scenario, all schools are equally effective, at least as far as their relationship to factor X is concerned, and no relative rankings exist. All values of u_S are identical to the adjusted school score, and both absolute rankings and relative rankings of schools produce the identical ordering of schools from best to worst. Neither panel A nor panel B represents the situation for adjusted school scores in Ontario.

A scenario closer to that which prevails in Ontario is represented in panel C, which shows a clear positive relationship between the value of factor X and schools' adjusted scores; in reality, of course, factor X represents many different variables that could affect those scores.

Indeed, although the school's socio-economic context plays an important role in its adjusted score, other factors — which we might think of as the "quality" of the school — are also important. Such factors are not observed directly. That is to say, schools that draw their students from similar socio-economic backgrounds still may produce different adjusted assessment results. Two schools are exactly comparable when the value of factor X is the same for both and when their predicted assessment scores are the same. But when the results for each school are calculated to take into account variations in the schools' socio-economic context as measured by factor X, one can obtain a relative ranking for each school. In panel C, that relative ranking places the school at a particular vertical distance from its predicted adjusted score, shown by the straight line in the panel. Thus, schools with a higher relative ranking lie above the line of predicted values, while those with lower relative rankings lie below the line. By way of example, four schools are labelled in the panel: schools 1, 2, and 4 are "good" schools, but school 3 is not as good — even though it and school 2 have the same value of factor X, school 2 is clearly a better school.

Panel D of Figure 6.2 presents the normal distribution of Z-scores for the schools in our sample. As noted earlier, any school whose Z-score is +1.29 or higher is "better" than 90 percent of schools. Schools 1, 2, and 4 all meet this criterion — that is, they all have the same Z-score. As a relative ranking of schools, the Z-scores represent a fair way to compare school effectiveness that takes into account variations in factor X across schools.

Ranking Schools: The Waterloo Region District School Board

In this section, I apply the methodology of relative rankings created by Z-scores to a sample of Ontario elementary schools, drawn from the Waterloo Region District School Board, and compare them with the absolute rankings used by, among others, the Fraser Institute. (Z-score ratings for all elementary schools in Ontario for which sufficient data exist can be found on the C.D. Howe Institute's web site at: www.cdhowe.org.)

As noted previously, the Waterloo board is of medium size and contains a wide variety of schools, both rich and poor, as well as urban, suburban, and rural. It thus represents a reasonable cross-section of the various types of schools found in Ontario, with one exception: no schools in the Waterloo board contain a significant aboriginal population. Table 6.1 lists the socio-economic characteristics of all 89 schools in the board and gives a sense of the variations in those characteristics that can occur across a specific set of schools.

Table 6.2 goes straight to the heart of the matter, showing both relative and absolute rankings (according to the Fraser Institute's methodology; see Cowley and Easton 2003) of schools in the Waterloo board, as well as rankings based on these schools' absolute assessment results. Both my relative Z-score rankings and the absolute rankings use the same sets of data from the 1998–99, 1999–2000, 2000–01, and 2001–02 academic years. However, the Fraser Institute includes only 65 of the board's 89 schools — and presents complete data for just 60 of them.

In its 2003 report, the Fraser Institute ranks 2,432 schools using a measure of performance between 0 and 10. For the schools in the Waterloo board, these rankings are found in column 12 of Table 6.2, "Overall Rating for the Past 4 Years." The details of the construction of the overall rating score are found in the Fraser report (Cowley and Easton 2003), but they may be summarized as follows.

The Fraser Institute's absolute ratings methodology begins by taking the EQAO Method 2 results — the assessment results excluding exempt students[8] — then, after taking into account variations in the average result by year and assessment, these results are averaged for each school. About 50 percent of the weight of the rating from 0 to 10 is based directly on these average scores. Another 30 percent is determined by a variable based on the percentage of exams written with sufficient material to grade that did not achieve either Level 3 or Level 4, which is a second way to measure average scores. The rating also includes the influence of a variable based on the difference between male and female results in the Grade 6 scores.

Figure 6.3 presents a visual picture of the absolute ratings of the 60 schools in Table 6.2 for which the Fraser Institute provides complete data. In the figure, each school's overall rating for the four academic years (column 12 in Table 6.2) is plotted against the average percentage of *all* students in both Grades 3 and 6 at a school scoring at either Level 3 or Level 4 (the average of the percentages found in columns 4 and 7 of Table 6.2). Figure 6.3 makes it clear that the Fraser Institute rating is also a measure of the percentage of *all* students at a school who achieve at Level 3 or higher. Indeed, the

---

[8] As noted earlier, this methodology is a potentially important weakness of the Fraser Institute's ratings, since it leaves the rankings process open to manipulation by schools that wish to do better by exempting students who would not have performed well on the assessments. Although this issue does not appear to be a problem at the present time, it could easily become one if Method 2 results continue to be used. As far as the Waterloo board is concerned, there is no evidence that exemption rates are being raised in order to achieve overall higher scores on tests actually written.

Table 6.1: *Socio-Economic Characteristics of Schools in the Waterloo Region District School Board*

School	Single-Parent Families	Official Language Spoken at Home	Aboriginal Status	Living in Canada < 5 years	Detached Homes	Average Household Income
	(%)	*(%)*	*(%)*	*(%)*	*(%)*	*(1995 $)*
A.R. Kaufman	33.2	81.7	1.4	13.3	25.1	46,874
Alison Park	27.0	88.1	0.6	3.9	52.9	54,260
Alpine	21.5	94.7	1.0	3.0	44.2	57,590
Avenue Road	23.1	86.1	0.4	4.5	37.1	53,888
Ayr	10.9	99.6	0.4	0.7	87.4	77,119
Baden	10.5	98.0	0.8	0.3	84.0	67,571
Blair Road	13.5	96.6	0.4	1.0	77.9	77,759
Breslau	11.9	93.6	0.8	1.6	84.5	75,259
Bridgeport	19.0	94.5	0.7	3.3	67.5	66,440
Brigadoon	20.1	97.0	0.4	1.6	64.0	77,151
Brighton	23.2	78.6	0.5	20.0	25.2	42,059
Cedar Creek	10.7	99.5	0.4	0.8	87.6	77,189
Cedarbrae	32.4	89.0	1.3	7.7	34.7	50,158
Centennial	25.1	98.1	0.3	1.0	62.9	59,402
Central	25.9	95.5	0.6	0.9	59.4	53,933
Chalmers Street	25.1	89.2	0.4	1.9	50.2	54,386
Clemens Mill	11.8	93.8	0.8	2.7	74.2	81,598
Conestogo	12.7	85.5	0.7	0.9	88.0	83,394
Coronation	26.8	95.8	0.7	2.0	50.8	53,914
Country Hill	20.4	90.4	1.4	5.1	59.1	63,255
Crestview	27.3	84.5	0.7	10.1	42.5	53,268
Dickson	28.0	97.6	1.2	0.6	47.0	54,655
Doon	18.2	96.5	0.4	1.5	65.6	73,845
Driftwood Park	10.7	95.3	0.2	1.4	90.3	75,058
Elgin Street	22.7	92.5	0.4	1.9	44.2	56,687
Elizabeth Ziegler	26.0	93.8	0.8	3.8	55.1	61,783
Empire	17.1	95.2	0.4	2.7	73.7	79,938
Floradale	8.1	69.7	0.0	2.5	81.6	64,758
Forest Glen	12.8	89.3	1.4	1.4	65.9	56,562
Forest Hill	29.6	89.0	1.0	6.3	41.8	51,545
Franklin	20.0	94.5	0.5	2.2	68.8	68,870
Glencairn	13.7	90.7	0.6	2.4	66.3	65,486
Grand View	29.8	97.6	1.0	1.3	49.6	55,588
Grandview	12.9	89.2	1.4	1.4	65.7	56,533
Harold W. Wagner	18.9	95.8	0.4	1.1	63.6	90,892
Heidelberg	7.7	84.1	0.2	1.0	79.4	81,952
Hespeler	12.7	96.6	0.4	0.9	82.8	74,873
Highland	22.8	95.9	0.5	1.1	60.3	69,104
Hillcrest	14.5	97.7	1.0	0.6	75.2	64,824
Howard Robertson	29.1	89.7	1.6	7.0	23.4	48,841
J.F. Carmichael	30.4	87.9	1.2	8.2	40.6	47,624
John Darling	12.1	93.2	0.3	2.4	82.5	73,283
John Mahood	14.5	96.8	0.2	0.2	69.2	70,088
Keatsway	23.0	95.3	0.9	4.3	51.7	83,614
King Edward	34.3	90.3	2.0	6.4	44.5	46,482

Table 6.1 - *continued*

Moved in Past Year	Moved in Past 5 Years	Unemployment Rate	Adults without High School Diploma	Adults with Some University	Total Enrollment
(%)	(%)	(%)	(%)	(%)	
23.6	58.7	14.8	48.2	20.5	720
16.6	42.1	10.2	61.6	11.3	219
18.4	55.2	4.5	44.5	16.5	469
15.7	50.1	8.0	60.3	10.6	287
9.3	38.1	1.8	39.5	23.0	319
9.7	33.5	2.0	47.8	20.7	215
9.6	41.1	4.9	47.0	19.1	272
7.6	47.0	2.8	41.3	22.2	341
14.2	52.2	5.4	41.3	23.5	391
16.6	45.8	4.7	36.9	24.8	452
31.3	67.4	7.1	24.7	58.6	154
9.4	37.8	2.0	39.5	23.2	412
25.3	56.2	12.1	45.5	25.0	381
14.7	41.4	5.0	52.3	15.2	251
15.3	42.9	7.5	56.0	13.9	369
15.9	45.9	8.7	60.1	10.1	398
11.0	52.0	4.6	42.5	23.4	693
10.2	29.9	1.6	49.5	24.0	119
22.1	53.6	8.4	57.7	9.9	363
15.3	44.9	4.8	47.6	17.2	308
18.5	50.5	9.9	49.8	18.8	667
19.2	48.5	5.7	51.3	14.4	124
14.0	40.9	4.4	39.2	22.6	372
8.3	30.4	2.7	38.0	25.2	487
13.4	51.9	7.4	54.2	11.6	390
21.6	52.5	3.5	41.9	30.7	365
13.9	40.5	3.3	30.4	43.1	467
6.8	21.3	0.6	70.5	10.4	173
11.4	34.7	1.8	44.1	17.2	581
18.6	51.2	9.5	47.5	17.7	611
13.2	43.0	4.9	42.8	23.1	563
13.1	49.7	4.3	47.0	16.5	543
15.5	46.9	5.0	54.5	12.4	215
11.3	34.6	1.8	44.0	17.1	171
14.4	43.2	4.3	38.0	33.0	334
8.5	26.8	2.5	54.0	18.2	72
11.8	38.6	3.5	45.6	20.5	472
15.9	46.3	5.4	46.7	17.7	557
14.9	49.2	4.8	46.3	17.2	194
21.6	54.9	7.5	52.5	14.1	624
18.5	52.8	8.7	48.5	19.5	601
9.7	36.9	2.4	37.1	25.6	431
9.8	33.0	1.4	47.5	19.7	291
16.3	43.8	7.6	25.6	50.8	418
20.5	54.3	7.8	49.6	19.6	371

Table 6.1 - *continued*

School	Single-Parent Families	Official Language Spoken at Home	Aboriginal Status	Living in Canada < 5 years	Detached Homes	Average Household Income
	(%)	*(%)*	*(%)*	*(%)*	*(%)*	*(1995 $)*
Laurelwood	14.5	92.4	0.3	5.1	70.1	91,613
Lexington	17.1	93.1	0.5	1.5	65.5	82,767
Lincoln Heights	21.5	96.5	1.3	3.5	57.9	71,220
Linwood	7.1	65.5	0.1	0.9	81.7	66,847
Mackenzie King	18.5	92.3	0.4	1.9	66.1	60,395
Manchester	27.1	87.4	0.4	1.3	55.4	50,105
Mary Johnston	17.0	93.6	0.2	3.7	66.5	116,158
Meadowlane	17.9	94.3	0.6	2.6	66.8	67,410
N.A. MacEachern	20.2	93.6	0.5	2.5	74.6	64,750
New Dundee	10.3	97.6	1.1	0.7	87.3	82,151
North Wilmot	14.0	97.7	0.5	0.2	85.9	67,526
Northlake Woods	8.8	96.1	0.4	1.8	73.4	79,713
Park Manor	14.4	94.9	0.2	0.5	69.8	69,234
Parkway	27.2	95.5	1.1	2.2	36.1	50,893
Pioneer Park	19.1	94.4	0.6	1.6	55.2	63,326
Preston	22.0	97.2	0.5	1.3	53.1	58,641
Prueter	30.5	96.1	1.2	1.7	63.1	50,078
Queen Elizabeth	36.1	84.1	1.6	5.2	33.1	41,599
Riverside	15.2	94.5	0.1	0.6	64.9	62,031
Rockway	33.5	89.4	0.9	6.4	36.0	47,525
Ryerson	16.4	95.8	0.3	1.8	79.4	63,966
Saginaw	7.2	93.0	0.6	3.2	80.0	97,312
Sandhills	15.5	88.0	0.3	5.0	62.0	69,335
Sandowne	17.2	95.5	1.1	1.5	69.4	73,602
Sheppard	26.4	93.1	1.2	1.8	65.4	63,222
Silverheights	15.5	97.5	0.6	0.7	64.8	66,227
Smithson	23.5	95.7	0.4	1.9	65.8	56,643
Southridge	26.3	89.2	0.4	5.5	45.0	58,688
St. Andrew's	20.0	96.6	0.6	0.9	68.4	69,192
St. Jacobs	11.0	81.9	1.6	2.3	71.1	75,329
Stanley Park	22.3	89.9	0.6	5.1	60.1	60,843
Suddaby	32.4	91.0	1.3	4.2	41.1	51,142
Tait Street	14.8	96.9	0.8	0.6	83.3	69,009
Three Bridge	8.5	77.5	1.2	2.1	72.4	74,925
Trillium	22.3	91.3	1.3	3.6	52.7	57,260
Wellesley	10.8	89.1	0.0	0.8	89.1	66,319
Westheights	11.9	94.3	0.2	1.9	85.2	74,401
Westmount	30.8	87.3	0.7	10.7	34.8	59,482
Westvale	12.2	95.1	0.5	1.8	86.0	78,074
Wilmot Senior	11.6	97.4	0.8	0.5	84.7	72,357
Wilson Avenue	35.3	90.7	1.5	6.6	33.2	47,348
Winston Churchill	34.6	89.4	0.5	8.0	18.8	48,834
Winterbourne	12.5	86.8	0.7	0.8	88.6	81,983
Woodland Park	11.8	97.0	0.3	0.6	74.6	79,981

Table 6.1 - *continued*

Moved in Past Year	Moved in Past 5 Years	Unemployment Rate	Adults without High School Diploma	Adults with Some University	Total Enrollment
(%)	(%)	(%)	(%)	(%)	
17.9	65.8	4.0	25.0	42.5	446
14.7	48.1	3.2	34.0	34.9	405
14.1	41.6	5.3	41.3	29.2	399
5.8	20.5	0.3	71.0	7.6	632
11.0	37.1	4.0	46.7	18.9	277
16.7	46.2	8.7	63.6	8.7	369
11.6	44.2	4.3	25.6	48.3	519
12.6	44.1	3.5	41.8	22.3	391
11.5	36.7	3.4	40.8	26.6	352
11.0	35.5	2.1	39.6	24.2	221
8.8	26.5	2.0	49.8	17.5	112
9.7	39.5	2.5	33.3	29.8	629
9.9	31.2	1.6	50.8	19.5	238
19.2	56.7	11.2	51.6	13.8	294
12.5	39.9	4.6	44.5	18.0	282
16.5	47.5	3.9	54.2	15.2	275
15.6	44.5	5.4	53.5	16.1	298
21.2	55.9	8.7	59.5	11.0	395
10.6	32.6	2.0	57.5	17.3	169
21.3	49.8	6.7	55.1	13.2	328
9.6	14.3	3.1	55.9	12.4	161
14.7	67.9	3.3	35.6	28.7	302
22.6	67.5	6.4	38.2	28.5	433
13.2	38.7	2.6	36.9	31.4	420
15.1	44.3	4.3	43.7	23.6	382
11.8	44.0	4.3	45.8	16.0	552
13.3	38.8	4.5	49.4	19.5	285
18.1	49.3	8.0	44.0	23.5	538
14.1	43.6	4.6	46.3	18.2	496
11.7	29.2	3.5	51.5	18.7	303
14.6	42.4	6.2	46.6	20.9	468
20.8	53.1	8.9	43.5	25.5	401
11.7	39.6	2.5	45.0	19.2	378
10.4	26.7	3.1	56.9	16.9	121
14.8	41.8	5.8	48.1	15.4	322
7.4	28.9	0.6	56.6	13.0	420
9.2	33.9	2.7	38.0	25.3	574
23.3	57.1	8.6	42.0	28.1	510
11.2	45.5	3.4	29.1	38.8	621
10.3	32.9	2.2	45.0	21.2	311
20.6	52.4	9.3	53.4	15.0	545
28.8	61.6	5.8	34.9	39.9	265
9.6	29.4	1.9	49.2	23.0	149
14.4	43.6	2.7	39.7	25.0	414

Table 6.2: *Measures Used to Rank Schools in the Waterloo Region District School Board*

					Grade 3	
School	Average Household Income *(1)*	Fraser Institute Rank by Quartile *(2)*	Z-score Rank by Quartile *(3)*	Students at Level 3 or Level 4 *(4)*	Z-score *(5)*	Z-score Rank in Region *(6)*
	(1995 $)			*(%)*		
Mary Johnston	116,158	1	2	70.6	0.10	18
Laurelwood	91,613	1	3	61.6	-0.17	46
Harold W. Wagner	90,892	2	3	48.3	-1.03	77
Keatsway	83,614	1	1	67.0	0.39	28
Lexington	82,767	1	1	66.5	0.56	21
Clemens Mill	81,598	1	1	61.8	0.42	26
Woodland Park	79,981	2	3	54.7	-0.23	49
Empire	79,948	1	1	82.3	1.69	4
Northlake Woods	79,713	1	2	53.6	-0.82	71
Westvale	78,074	2	3	56.3	-0.79	68
Blair Road	77,759	1	1	60.8	0.46	23
Breslau	75,259	1	1	80.1	1.89	1
Hespeler	74,873	1	1	66.9	1.01	10
Sandowne	73,602	1	2	61.6	0.27	31
Lincoln Heights	71,220	3	4	47.1	-0.70	65
Tait Street	69,009	2	2	49.6	-0.50	57
Franklin	68,870	2	2	51.3	-0.27	51
Meadowlane	67,410	1	4	52.3	-0.20	48
Linwood	66,847	1	4	44.0	-1.10	80
Bridgeport	66,440	1	1	64.5	1.02	9
Wellesley	66,319	2	2	50.3	-0.54	60
Silverheights	66,227	1	4	46.2	-0.51	58
Glencairn	65,486	3	4	37.0	-1.19	82
Hillcrest	64,823	2	2	58.3	0.70	17
N.A. MacEachern	64,750	1	1	53.4	-0.15	52
Ryerson	63,966	1	1	64.2	1.13	5
Country Hills	63,254	2	2	54.8	0.44	25
Sheppard	63,222	2	1	60.3	0.80	15
Elizabeth Ziegler	61,783	1	1	57.0	0.60	20
Mackenzie King	60,395	1	3	54.4	0.10	37
Westmount	59,482	2	2	36.2	-1.03	78
Centennial	59,402	3	1	38.7	-0.79	69
Southridge	58,688	2	1	58.7	0.93	13
Preston	58,640	2	1	53.9	0.79	16
Alpine	57,590	4	4	39.3	-0.62	62
Trillium	57,260	4	4	41.9	-0.56	61
Elgin Street	56,687	3	1	34.5	-0.94	75
Smithson	56,643	2	2	53.3	0.29	30
Grand View	55,588	4	4	36.2	-0.66	63
Forest Glen	56,562	3	4	48.6	-0.46	56
Stewart Avenue	55,439	1	2	47.3	0.08	39
Alison Park	54,260	4	3	41.0	-0.11	45
Central	53,933	3	3	36.3	-0.73	67
Coronation	53,914	4	4	29.4	-0.82	72
Avenue Road	53,888	3	2	30.3	-1.14	81

Table 6.2 - *continued*

Grade 6					Fraser Institute		
Students at Level 3 or Level 4 *(7)* (%)	Z-score *(8)*	Z-score Rank in Region *(9)*	Average Z-score Grades 3 & 6 *(10)*	Ranking in Province Past 4 Years *(11)*	Overall Rating Past 4 Years *(12)*	Overall Rating 2001-02 *(13)*	Fraser Rank in Region *(14)*
71.8	0.02	37	0.06	120	8.6	8.9	2
65.3	-0.01	38	-0.09	343	7.7	8.5	6
68.7	0.71	16	-0.16	599	7.0	7.2	16
61.2	-0.58	63	-0.09	377	7.6	8.0	8
70.2	0.81	12	0.68	264	8.0	7.4	4
67.3	1.20	8	0.81	513	7.2	7.5	11
54.3	-0.30	47	0.27	905	6.5	7.1	20
79.8	1.28	6	1.49	41	9.3	8.9	1
69.1	0.78	14	-0.02	288	7.9	8.2	5
65.6	0.09	33	-0.35	707	6.8	6.7	17
72.8	1.69	1	1.08	406	7.5	6.4	10
54.9	-0.26	46	0.81	156	8.4	8.3	3
60.7	0.56	20	0.79	555	7.1	7.2	14
59.0	-0.05	39	0.33	513	7.2	7.2	12
51.3	-0.42	53	-0.56	1,437	5.6	5.7	32
58.1	0.42	24	-0.04	1,049	6.1	6.0	24
56.8	0.42	26	0.07	1,094	6.2	5.4	26
41.1	-1.18	74	-0.69	1,542	5.4	4.7	34
43.5	-0.71	67	-0.90	1,772	5.0	4.7	41
64.3	1.10	11	1.06	513	7.2	8.1	13
61.6	0.78	13	0.12	905	6.5	6.4	21
43.2	-0.57	61	-0.54	1,658	5.2	5.3	37
51.7	0.21	29	-0.59	1,720	5.1	5.3	40
49.2	-0.18	41	0.26	769	6.7	7.2	18
70.8	1.33	5	0.49	555	7.1	7.4	15
68.3	1.64	2	1.39	377	7.6	7.8	9
44.1	-0.47	56	-0.02	1,273	5.9	6.1	28
56.4	0.49	21	0.64	905	6.5	5.8	22
65.8	1.20	9	0.90	343	7.7	6.2	7
46.0	-0.61	64	-0.26	1,437	5.6	5.4	33
64.2	1.54	3	0.25	1,170	6.1	6.2	27
48.1	0.10	32	0.34	1,824	4.9	5.2	44
55.3	0.72	15	0.82	769	6.7	6.8	19
49.0	0.29	28	0.54	1,326	5.8	5.5	30
38.5	-0.62	65	-0.62	2,094	4.1	4.9	52
40.8	-0.57	62	-0.56	1,920	4.7	5.5	46
47.2	0.42	25	-0.26	1,824	4.9	5.5	45
56.1	0.59	18	0.44	905	6.5	6.7	23
38.5	-0.43	54	-0.54	1,964	4.6	5.5	49
42.0	-0.83	68	-0.65	1,658	5.2	5.4	38
41.3	-0.15	40	-0.03	1,658	5.2	5.4	39
39.3	-0.26	45	-0.18	1,920	4.7	5.8	47
45.2	0.16	30	-0.28	1,599	5.3	4.7	35
24.0	-1.37	76	-1.10	2,371	3.0	3.5	58
54.1	1.17	10	0.01	1,772	5.0	5.8	42

Table 6.2: *Measures Used to Rank Schools in the Waterloo Region District School Board*

				Grade 3		
School	Average Household Income (1)	Fraser Institute Rank by Quartile (2)	Z-score Rank by Quartile (3)	Students at Level 3 or Level 4 (4)	Z-score (5)	Z-score Rank in Region (6)
	(1995 $)			(%)		
Crestview	53,268	4	4	36.0	-0.90	74
Forest Hill	51,545	2	1	48.9	0.46	24
Suddaby	51,142	3	2	46.4	0.14	35
Parkway	50,893	4	2	50.9	1.06	8
Cedarbrae	50,158	4	3	41.5	0.00	40
Manchester	50,105	4	4	24.3	-1.52	83
Prueter	50,078	3	3	46.3	0.16	34
Howard Robertson	48,841	4	4	34.9	-0.40	53
Winston Churchill	48,834	4	4	35.3	-1.04	79
J.F. Carmichael	47,624	2	2	51.4	0.67	18
Rockway	47,525	3	2	39.1	-0.06	43
Wilson Avenue	47,348	4	3	34.6	-0.41	55
A.R. Kaufman	46,873	4	3	32.5	-0.71	66
King Edward	46,482	4	4	33.2	-0.81	70
Queen Elizabeth	43,599	4	4	27.4	-0.84	73
Ayr	77,119	—	—	70.3	0.96	12
Chalmers Street	54,386	—	—	40.2	-0.19	47
Floradale	64,758	—	—	68.0	1.11	6
Grandview	56,533	—	—	63.0	0.84	14
St. Jacobs	75,329	—	—	58.3	0.21	32
Schools Not Included in the Fraser Institute Rankings						
Baden	67,570			56.4	-0.01	41
Brigadoon	77,151			65.0	0.99	11
Brighton	42,059			75.1	1.75	3
Cedar Creek	77,189			57.8	-0.26	50
Conestogo	83,394					
Dickson	54,655			34.2	-0.69	64
Doon	73,835					
Driftwood Park	75,057			66.5	0.41	27
Heidelberg	81,952			69.7	1.08	7
Highland	69,104			51.3	0.20	33
John Darling	73,283			55.5	-0.53	59
John Mahood	70,088			51.4	0.41	54
New Dundee	82,151			59.5	-0.04	42
North Wilmot	67,526			60.9	0.51	22
Park Manor	69,234					
Pioneer Park	63,326			40.8	-1.01	76
Riverside	62,031			70.9	1.76	2
Saginaw	97,332			60.3	-0.09	44
Sandhills	69,335			54.5	0.11	36
St. Andrew's	69,192					
Three Bridge	74,925			62.9	0.65	19
Westheights	74,401					
Wilmot Senior	72,357					
Winterbourne	81,983			63.4	0.29	29

Table 6.2 - *continued*

Grade 6					Fraser Institute		
Students at Level 3 or Level 4 *(7)*	Z-score *(8)*	Z-score Rank in Region *(9)*	Average Z-score Grades 3 & 6 *(10)*	Ranking in Province Past 4 Years *(11)*	Overall Rating Past 4 Years *(12)*	Overall Rating 2001–02 *(13)*	Fraser Rank in Region *(14)*
(%)							
38.8	-0.43	55	-0.66	1,920	4.7	3.8	48
48.5	0.64	17	0.55	1,049	6.3	6.1	25
49.2	0.33	27	0.23	1,326	5.8	6.0	31
36.4	-0.31	49	0.38	2,094	4.3	2.9	53
38.1	-0.51	59	-0.25	1,964	4.6	5.3	50
35.6	-0.36	50	-0.94	2,387	2.7	2.4	59
40.2	-0.38	51	-0.11	1,772	5.0	5.6	43
24.9	-1.23	75	-0.82	2,314	3.4	4.1	57
47.2	-0.48	58	-0.76	1,964	4.6	5.5	51
39.9	-0.22	41	0.21	1,273	5.9	6.5	29
40.8	0.13	31	0.04	1,599	5.3	3.6	36
33.3	-0.41	52	-0.41	2,094	4.3	3.9	54
43.3	0.47	22	-0.12	2,193	4.0	4.2	55
30.4	-1.09	73	-0.95	2,213	3.2	3.9	56
18.4	-1.60	77	-1.22	2,400	2.4	2.6	60
60.8	0.06	34	0.51	n.a.	n.a.	7.8	n.a.
56.7	1.40	4	0.61	n.a.	n.a.	6.7	n.a.
67.3	1.21	7	1.16	n.a.	n.a.	8.0	n.a.
47.4	-0.66	66	0.09	n.a.	n.a.	5.2	n.a.
46.1	-0.90	71	-0.34	n.a.	n.a.	6.4	n.a.
Schools Not Included in the Fraser Institute Rankings							
			-0.01				
			0.99				
56.8	-0.56	60	0.60				
56.7	-0.30	48	-0.28				
52.7	-0.83	69	-0.83				
			-0.69				
52.5	-0.18	42	-0.18				
			0.41				
55.0	-0.22	44	0.43				
			0.20				
			-0.53				
			-0.41				
			-0.04				
			0.51				
46.3	-0.89	70	-0.89				
			-1.01				
			1.76				
57.3	-0.47	57	-0.28				
56.0	0.03	36	0.07				
38.9	-1.03	72	-1.03				
61.9	0.56	19	0.60				
59.3	0.04	15	0.04				
61.7	0.43	23	0.43				
			0.29				

Figure 6.3: ***The Relationship between Fraser Institute Ratings and Students Achieving Assessment Levels 3 and 4 in the Waterloo Region District School Board***

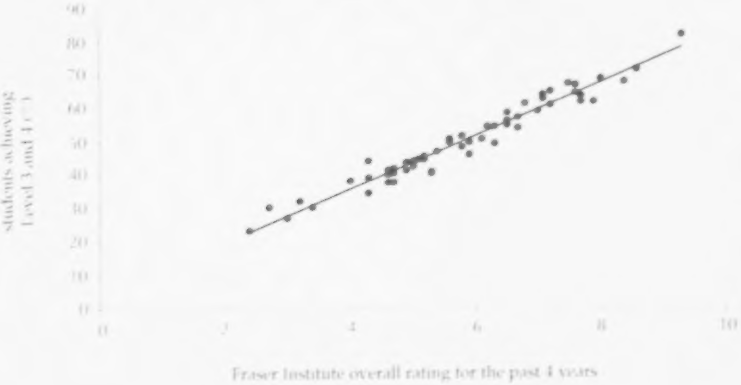


figure bears a striking resemblance to Figure 6.2, panel A, which displays a regression with an R^2 equal to 1, implying a very tight relationship between adjusted school scores and variable X. In effect, the Fraser Institute's overall rating number is, for all intents and purposes, simply a measure of the percentage of all students at the school, averaged over all the assessments, that achieves at Level 3 or Level 4 — in other words, it is an absolute achievement measure.

As an aside, it is interesting that the absolute rating measure also shows large year-over-year variability at the same school. The 2003 Fraser Institute report card includes a comparably constructed numerical rating variable, scaled between 0 and 10, that uses only the results from the 2001–02 academic year (column 13 in Table 6.2). For the schools in the table, the average difference between the overall rating for the four years and the rating for 2001–02 (columns 12 and 13) is –0.05, with a standard deviation of 0.61, a maximum of 1.7, and a minimum of –1.1. These are fairly large changes. About half the schools "improve" and half get "worse" in 2001–02 relative to the overall four-year performance. The insight here is that classes at schools — even averaging over both Grade 3 and Grade 6 — vary quite a lot from year to year; indeed, the Fraser

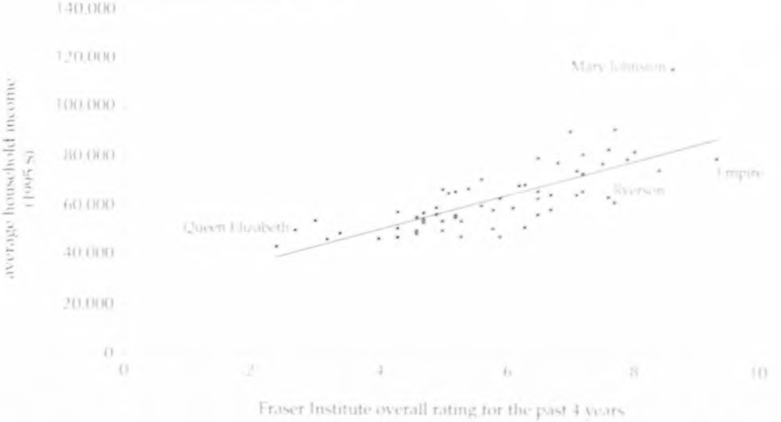
Institute results confirm my earlier conclusion that looking at year-over-year variations is not likely to be a very useful assessment of changes in school quality.

Column 11 in Table 6.2 is the Fraser Institute's ranking of the 60 schools in the Waterloo board sample out of the 2,432 schools it ranks for the entire province. For convenience, column 14 ranks the 60 schools in the sample from best to worst (first to sixtieth). Thus, for example, Mary Johnston School ranks 120th in the province and second among the 60 Waterloo board schools in the sample. Note, however, that Table 6.2 is *not* organized according to the Fraser Institute's rankings; rather, the 60 schools are ranked by average household income, which should help the reader to gain further insight into the difference between absolute rankings and relative Z-score rankings.

If we group the absolute rankings of the 60 schools into quartiles — that is, in groups of 25 percent of the total number of schools, which, in this case, is groups of 15 schools — shown in column 2 of Table 6.2, we see that, of the 15 schools with the highest average household income, 11 are also in the highest quartile of the absolute rankings. In other words, a high-income school in the Waterloo region is also very likely to be highly ranked on the absolute scale. Similarly, 9 of the 15 schools in the lowest income quartile are also in the lowest quartile of the absolute rankings, and not one is in the first quartile. Thus, a low-income school in the Waterloo region is also very likely to rank poorly on the absolute scale. This suggests, in short, that the Fraser Institute's overall rating (column 12) is in large part a ranking of schools by household income.

Figure 6.4 confirms that analysis. The visual connection between the overall rating from the four-year period and household income is obvious. The straight line in the figure is the estimated regression relationship between the overall rating and income. The R^2 on this regression (reported in Table A6.2 in the technical appendix) is 0.59, which is a very high value compared with the R^2 in other regressions we have seen. On the other hand, the fact that the R^2 is certainly not 1 (as was the case in Figure 6.3) means there is some information in the overall rating that is not already contained in the

Figure 6.4: ***The Relationship between Fraser Institute Ratings and Average Household Income in the Waterloo Region District School Board***



income measure, although the information is quite limited. In Figure 6.4, Ryerson and Empire show as schools that perform well, given their level of household income. On the other hand, the figure also makes it clear that creating a ranking that compares Queen Elizabeth and Mary Johnston schools is a completely uninteresting exercise. The overall rating, at least between schools at substantially different income levels, simply reflects whatever factors that drive the differences in income levels that are also associated with strong assessment results, not factors that are part of the educational process as implemented at the school. Is there a better way to produce a ranking of schools? Are the relative, Z-scores of schools discussed earlier in this chapter more useful?

I believe that the answer to this important question is yes. In Table 6.2, columns 5 and 8 show the Z-scores for Grades 3 and 6 for the 60 schools in the Waterloo board sample, and the average of the two Z-scores appears in column 10. Think of the average Z-score across the two grades as a way to rank a school combining both grades. A high score indicates a school that is better than other schools in the rest of the province after taking into account the school's socio-economic setting using a variety of indicators.

If we group the relative rankings of the 60 schools into quartiles (column 3), we can compare the Z-score ranking and the Fraser Institute ranking. We then see that, of the 15 schools with the highest average household income, 6 are in the first quartile of Z-scores, 3 are in the second, 5 are in the third, and 1 is in the last quartile. It is immediately apparent that the relationship of the average Z-score to income is much weaker than that of the absolute ranking to income, which implies, in turn, that the Z-scores contain substantially more information about schools that perform well given their socio-economic context.

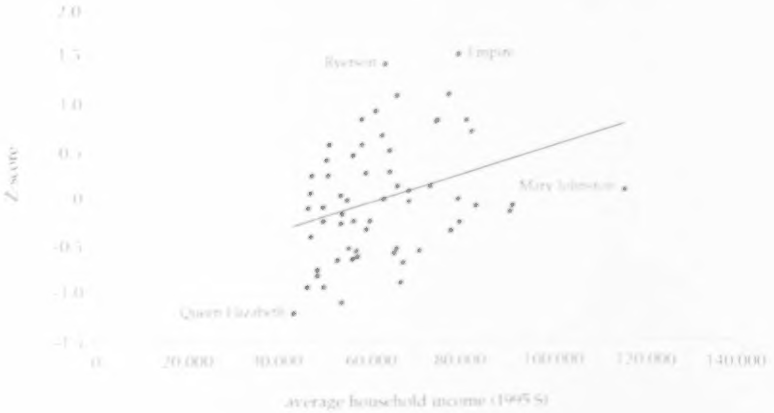
Figure 6.5 plots the average Z-score against household income for the 60 Waterloo schools. The estimated regression line (the straight line) has an R^2 of 0.10 (reported in Table A6.2 in the technical appendix), which indicates the weak relationship between a high Z-score and high average household income.[9] Recall that a high Z-score identifies a school that outperforms a school with comparable socio-economic characteristics. Thus, in this analysis, while Empire continues to outperform similar schools, Mary Johnston, although it ranks second in the region and very highly in the province according to the absolute scale, does not outperform schools in the rest of the province with similar socio-economic characteristics. Ryerson ranks highly on the absolute scale, but does even better on the relative scale — indeed, with an average Z-score of 1.39, Ryerson is better than 92 percent of schools in the province.

It is important to realize that both ranking systems have some ability to detect good schools. Empire and Ryerson show up as good

---

[9] Statisticians among the readers will realize that the predicted association between the Z-score and income is zero since the Z-score is constructed as a residual from a regression that includes household income. If the Z-score had been constructed using only data from the Waterloo board, the association between the Z-score and income would be zero, as would be the slope of the regression line. The regression result for the 60 schools for which the Z-score is related to income is a small-sample phenomenon because the Z-scores use all the data from all schools in the province.

Figure 6.5: ***The Relationship between Z-scores and Average Household Income in the Waterloo Region District School Board***

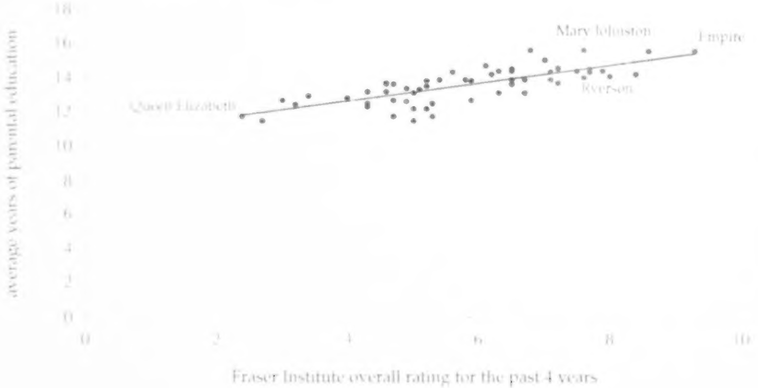


schools in both rankings, but the relative rankings show just how good a school Ryerson is compared with similar schools in the province. The relative rankings can, of course, also detect schools that perform poorly compared with other, similar schools. Queen Elizabeth is a low-income school that also ranks very low on the absolute scale of assessment performance, but the fact that it is a low-income school does not necessarily account for its poor performance, as the Fraser Institute rankings might suggest. Indeed, Queen Elizabeth scores much worse than schools with similar socioeconomic characteristics in the rest of the province, and a reasonable person would ask why.

In short, rankings based on absolute scores do reveal some information about the quality of education, and they are capable of capturing extremes in the performances of students at schools, both good and bad. However, they do not generate a fair or even interesting comparison of very different schools, nor can they easily pick out better performances by similar schools.

Those who use absolute rankings do recognize the need to compare similar schools, and the Fraser Institute report includes two methods

Figure 6.6: ***The Relationship between Fraser Institute Ratings and Parental Educational Achievement in the Waterloo Region District School Board***



that offer the possibility of an adjustment for context.[10] It is fair to say, however, that neither method receives much emphasis in the Fraser Institute report. The only school context variable in the Fraser Institute's analysis is a very carefully constructed variable that uses postal codes and census data for individual students tied to that postal code to measure the average number of years of education of the most educated parent in a household with children of elementary school age. This is, in fact, a more precise measure of the education level of adults in the postal codes in which the school's students live than the one I used in Chapter 3.[11]

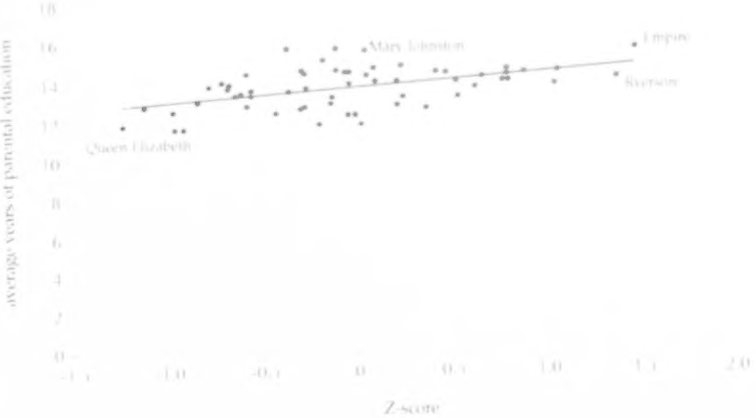
Using this methodology, comparable schools are defined as those with the same value of the parental education variable, and it is an excellent variable to use. As Figure 6.6 shows, even in a very small sample of only 60 schools, this variable works quite well in placing

---

[10] One of these methods is discussed above; the other is discussed at some length in the technical appendix to this chapter.

[11] The difference is that, for the Fraser Institute study, Statistics Canada worked with the individual responses to the census rather than aggregating the responses to the smallest geographic unit.

Figure 6.7: *The Relationship between Z-scores and Parental Education in the Waterloo Region District School Board*



school results in context. We see that Empire and Mary Johnston have similar levels of parental education, but Empire produces much higher assessment results and, therefore, can be considered a "better" school. Ryerson also performs better than schools with similar levels of parental education even within its own board. Once again, Queen Elizabeth underperforms relative to other schools in the Waterloo board with identical levels of parental education.

It is certainly possible, using the parental education variable, to create relative rankings and place school comparisons in context, but the Fraser Institute chose not to undertake this task. As Figure 6.7 shows, however, the Z-score methodology is superior in that it makes adjustments for context that are more convincing by incorporating information — such as variations in language, immigration, mobility, aboriginal status, housing status, and income — that go well beyond average years of parental education. Certainly, the outlying nature of Empire and Ryerson schools is even easier to see in Figure 6.7. In addition, we have seen that the effect of parents' educational status on students' assessment results appears to differ between Grade 3 and Grade 6. The use of average Z-scores allows this effect to be included in the analysis.

Conclusions

Rankings that are not appropriately adjusted for schools' socio-economic characteristics reveal very little about the performance of educators. Rankings that depend on the absolute level of schools' performance capture only the very strongest and very weakest performances, and even then not in all circumstances. In fact, to be blunt, my view is that absolute rankings that use results from schools drawing from vastly different socio-economic communities are virtually meaningless.

A teacher who transferred from a high-income school to a low-income school in the Waterloo region remarked, as I was describing this project: "I went from being one of the best teachers in the region to being one of the worst teachers in the region in one year." The statement, made in jest, is obviously not true — the teacher likely did not change anything that he did — but it reveals the problems in using absolute rankings, which undermines the valid use of the results of the assessment process.

I believe that relative rankings based on the performance of schools in similar socio-economic settings are much more appropriate and useful. Then, with such comparisons in hand, we need to find out what makes a school outperform similar schools and why that occurs. Some of the answers to that question become apparent in the next chapter, which relates the contents of interviews I conducted with supervisory and teaching staff at schools that clearly outperformed similar schools.

Chapter Summary

This chapter explores the use of EQAO assessment results to rank schools — that is, to identify schools that perform better than other schools in communities with similar socio-economic characteristics. Two methods of ranking schools are presented. The first, called an "absolute" ranking, ranks schools according to the level of their assessment results — the larger the percentage of students that achieves at Level 3 or Level 4, the higher the school's ranking — and is the method used by the Fraser Institute in its studies of Ontario elementary schools.

The second method, which I strongly prefer, is to create a "relative" ranking of schools. Using this method, I explain how to identify two or more schools that are predicted to have the same assessment result based on the socio-economic characteristics of their school communities. Schools are then ranked according to whether they produce a better result or worse than a comparable school. In a relative ranking, all schools are compared only after taking into account variations in their socio-economic context. A relative ranking recognizes that it is not very interesting to make an absolute comparison of assessment results of schools that draw their students from dissimilar communities. For example, comparing the assessment results of students at a low-income school where most parents have no postsecondary education and those at a high-income school where virtually all parents have a university degree tells us little about the relative performance of the two schools. However, a relative ranking allows us to determine how a particular school performs compared with other schools with a similar socio-economic setting.

The process of creating a relative ranking starts with determining the percentage of all students in Grade 3 or Grade 6 at each school that achieves at Level 3 or Level 4 on one of 24 standardized assessments given over the four academic years 1998–99 to 2001–02. In Chapter 4, I established that the average value of the assessment result — whether on the same assessment in different years or on different assessments in the same year — varies by year and type of assessment. In comparing schools across years, the mean of each assessment result (by type and academic year) is subtracted from the actual assessment result to create an adjusted assessment result. Thus, a school could have up to 12 adjusted assessment results in each grade. These 12 adjusted assessment results are then averaged separately for each school. Two single numerical values, one for Grade 3 and one for Grade 6, now represent each school's performance over four years on the Grade 3 and Grade 6 assessments. I call these values "adjusted school scores." A value of +10 indicates that 10 percentage points more students in each year at that school achieved at Level 3 or Level 4 on assessments than at the average school in the province in that grade.

The next step in constructing a relative ranking is to measure the statistical relationship between various indicators of the school community's socio-economic context and the adjusted school scores. Using the results from Chapter 4, I undertake the measurement separately for each grade. The variables describing the school community include the percentage of single-parent households, the linguistic background of the school community, its immigrant composition, the logarithm of average household income, mobility within the school community, and the educational attainment of adults in the school community. Variations in school community context variables are associated with 40 percent of the variation in Grade 3 school scores and 44 percent of the variation in Grade 6 school scores. The remaining percentages of the variations in school results (60 percent for Grade 3 and 56 percent for Grade 6) cannot be accounted for by variations in the socio-economic context in which schools operate, and it is this remaining variation that is the basis for the relative ranking of schools.

To see how this methodology works, consider the following example. Imagine that school communities vary only by average household income, and that average household income at school A is $10,000 larger than at school B, while average household income at school C is the same as at school A. Using the relationship estimated for all schools in the province, school A's adjusted score is predicted to be 2 percentage points higher than school B's, while the adjusted scores of schools A and C are predicted to be the same. The predictions in this example are based simply on the knowledge that incomes are $10,000 higher at school A than at school B and that incomes are the same at schools A and C. In the chapter predictions, however, many socio-economic factors other than income are used to predict schools' adjusted scores. If school A's adjusted score is 5 percentage points higher than school B's, then school A is "better" than school B by 3 percentage points. If school C's adjusted score is 3 percentage points higher than school A's, then school C is "better" than school A by 3 percentage points, while schools A and C are better schools than school B by the same amount. It is then very useful to be able to assess just how much better (or worse) schools B and C are than school A. This scaling is carried out by giving each school a Z-score for each grade. A school Z-score is a value on the normal distribution. A Z-score of +1.29 indicates that a school is better than 90 percent of other schools after adjusting for the socio-economic context in which a school operates. A Z-score of −1.29 indicates that 90 percent of other schools outperform this school after adjusting for the socio-economic context in which a school operates. The Z-score of a school in a grade is thus a relative ranking of that school.

To illustrate the difference between absolute rankings and relative rankings, I apply the two types of rankings to a sample of schools in the Waterloo Region District School Board. I explain the rankings used by the Fraser Institute and show that they are mainly a transformation of absolute school scores — a school with a higher percentage of students performing at Level 3 or Level 4 receives a higher rating. Moreover, there is an obvious and close visual and statistical connection between household income levels in school communities and the school's position in the Fraser Institute rankings — the higher the income of the school community, the higher the school's absolute ranking. In short, the absolute rankings used by the Fraser Institute reveal little beyond the socio-economic characteristics of the school community.

Much more revealing are relative rankings based on school Z-scores. Thus, for example, Mary Johnston, a high-income school in the Waterloo board, ranks second among schools in that board and 120th out of 2,432 schools in the province according to the absolute scale, but receives a Z-score of zero because its performance is only average among Ontario elementary schools with similar socio-economic community characteristics. In contrast, Ryerson, a middle-income school in the Waterloo board, ranks a respectable 377th in the province on the absolute scale, but according to the relative ranking its Z-score suggests that performs better than 90 percent of Ontario schools with similar socio-economic community characteristics. Low-income Queen Elizabeth School, ranked last in the Waterloo region and 2,400th in the province on the absolute scale, performs even worse than comparable schools according the relative rankings.

In summary, although rankings based on absolute scores contain some useful information, relative rankings contain much more useful information and are a much more valuable tool for assessing the performance of Ontario elementary schools.

Technical Appendix

This chapter uses regression techniques already explained in the appendix to Chapter 4. The equation estimated over the sample of schools is:

$$S_S = b_0 + b_1 X^1{}_S + b_2 X^2{}_S + b_3 X^3{}_S + b_4 X^4{}_S + u_S . \qquad \text{(A6.1)}$$

The various X variables found in the left-hand column of Table A6.1 are associated with the coefficient estimates of b_1, b_2, b_3, and so on. The estimation and interpretation of such coefficients have already been discussed in previous chapters and need not be repeated here.

The new concepts in Chapter 6 involve estimating and interpreting the values of u_S as "residuals" from equation (6.1). A calculation defines, for each school, a "fitted value," denoted S_S and equal to

$$S_S = b_0 + b_1 X^1{}_S + b_2 X^2{}_S + b_3 X^3{}_S + b_4 X^4{}_S , \qquad \text{(A6.2)}$$

where the coefficients b_0, b_1, b_2, b_3, b_4... are estimates of the coefficients listed in Table A6.1. The fitted values are the adjusted school scores predicted by a particular school's values of the whole group of X variables and the estimated coefficients (b_0, b_1, b_2, b_3, b_4...) found in Table A6.1. If a school community has a higher income, the school is predicted to have a higher adjusted score. If a school community has more single parents, the school is predicted to have a lower adjusted score. Different schools, with different combinations

Table A6.1: ***The Explanation of Adjusted School Scores***

Coefficient on School Context Variable	Grade 3 (Standard Error)	Grade 6 (Standard Error)
Single-parent families (%)	–0.10	–0.18
	(0.04)	(0.04)
Official language spoken at home (%)	–0.13	–0.05
	(0.04)	(0.04)
Aboriginal descent (%)	–0.58	–0.46
	(0.07)	(0.08)
Living in Canada < 5 years (%)	0.17	0.12
	(0.07)	(0.07)
Living in detached home (%)	0.06	0.04
	(0.01)	(0.01)
Logarithm of household income	7.08	7.06
	(1.11)	(1.11)
Moved in past year (%)	–0.46	–0.21
	(0.09)	(0.10)
Moved in past 5 years (%)	–0.04	–0.07
	(0.04)	(0.04)
Unemployment rate (%)	–0.28	–0.27
	(0.09)	(0.09)
Adults without high school diploma (%)	–0.20	–0.07
	(0.04)	(0.05)
Adults with some university (%)	0.26	0.50
	(0.04)	(0.05)
Total enrollment	–0.0007	–0.006
	(0.001)	(0.0001)
Number of schools	3,583	3,402
R^2	0.40	0.44

Note: All coefficients are significantly different from zero except those shaded. The variable descriptions are found in Chapter 3.

of X variables — that is, with different socio-economic settings — could have a lower, higher, or the same predicted score. With many schools in the data set, there are many predictions of school scores and thus many schools with the same predicted scores. Figure A6.1 plots, for every school in the province, a point whose coordinates are the predicted adjusted score and the actual adjusted score. There is one plot for Grade 3 results and one plot for Grade 6 results. Two observations can be made. First, the relationship between actual results and predicted results is clearly positive.[12] Second, because of the large number of schools, we have every reason to believe there are sufficient data to make good predictions for adjusted school scores based on socio-economic factors. However, most of the focus of the chapter is on the difference between the actual adjusted school score and the fitted value.

These differences, or regression residuals (to use more technical language), are calculated as

$$u_S = S_S - \mathrm{S}_S = S_S - [\, b_0 + b_1 X^1{}_S + b_2 X^2{}_S + b_3 X^3{}_S + b_4 X^4{}_S \,]. \qquad \text{(A6.3)}$$

If u_S is zero, then the school's performance in that grade is exactly as predicted by the combination of variables that describe its socio-economic context, and the school sits on the 45-degree line drawn through Figure A6.1. If u_S is positive, then the school performs better than other schools after taking into account variations in its socio-economic context and provincewide variations in the average results on different assessments written at that school; the school sits above the 45-degree lines passing through Figure A6.1.

One final step is needed to interpret these values of u_S. The units of u_S in equation (A6.3) are percentages of students at a school — that is, a value of 10 indicates that, in this school, 10 percentage points more students achieve at Level 3 or Level 4 than were predicted to achieve at Level 3 or Level 4 at the average school in the province. In the text, I established that a value of $u_S = 10$ is a substantial variation in education outcomes. The final step in the construction of relative school rankings is to ask whether it is common or uncommon to observe values of u_S approximately equal to 10.

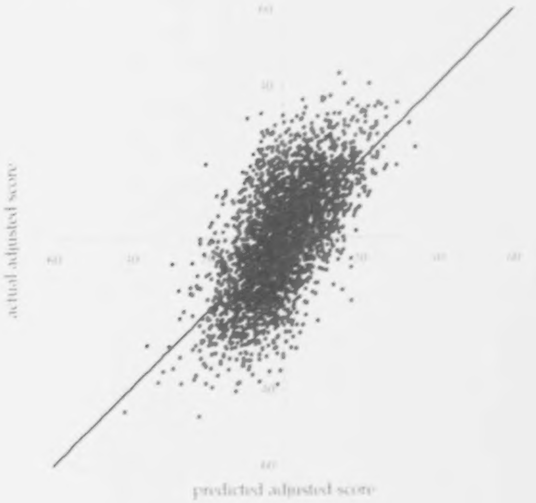
The most common statistical method of comparing variables is to take values of a variable and create a "normalized variable." Figure 6.1 presented visual pictures of histograms of the residuals,

---

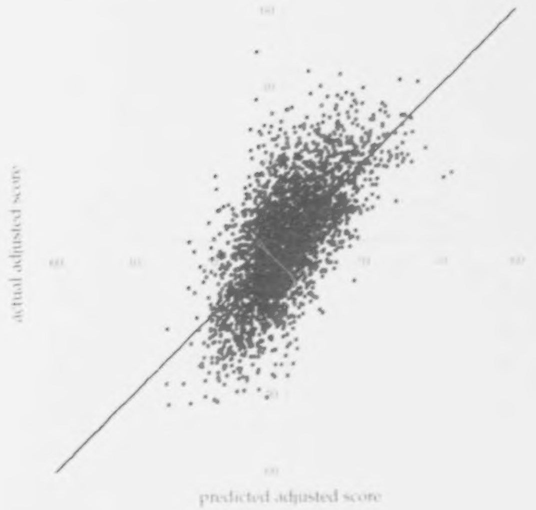
[12] This is associated with the fairly large value of the R^2. If the R^2 were 1.0, as in Figure 6.2, panel A, then all of the points in this diagram would lie on a 45-degree line from the origin. All predicted values would be actual values and all values of u_S would be zero. If schools behaved as in Figure 6.2, panel B, then the actual values would be randomly scattered relative to the predicted values.

Figure A6.1: ***The Relationship between Predicted Adjusted Scores and Actual Adjusted Scores, Grades 3 and 6***

Panel A: Grade 3 scores



Panel B: Grade 6 scores



the values of u_S, from the Grade 3 and Grade 6 equation, respectively. Figure 6.1 looks like a "normal" distribution. The fourth equation created a standardized value of the residual where

$$Z_S = u_S \,/\, (\text{standard deviation of } u_S)\,. \tag{A6.4}$$

One statistical issue in equation (A6.4) is that an estimate of the standard deviation of u_S is needed. There are three possibilities; fortunately, however, the results are identical to the third decimal point in all three. In the first case, we directly use the standardized residuals from the regression package. These ignore the position (the value of the X variables) at which this school is situated. In the second case, we take into account where the prediction of the school's score is taking place. If a school's X values are very far from the mean value of the X variables, then the estimate of the standard error is slightly larger. This is known as the "standard error of the forecast" and reflects the fact that the estimates of the values of the *b* coefficients are most precise at the large mass of observations that are near the average school. In the third case, we allow the error terms, u_S, to be heteroscedastic and then correct for this effect. As already noted, the Z-scores for the schools are invariant to the choice of normalization method.

Table A6.2 presents the regression results from Figures 6.3 through 6.7. In each figure, a regression line shows the estimated relationship between the variable on the vertical axis (the dependent or left-hand variable) and the variable on the horizontal axis (the independent or right-hand variable). The slope of that line is reported in Table A6.2 as well as the R^2 from the regression.

The Fraser Institute's Attempt to Adjust for Context

As noted in the text, the Fraser Institute report (Cowley and Easton 2003) makes a second attempt to look at school context. It estimates four equations, each of which links, at a school, the Grade 6 reading result to the Grade 3 reading result from the "same" class — that is,

Table A6.2: ***Regressions Estimated in Figures 6.3, 6.4, 6.5, 6.6, and 6.7***

	Dependent (left-hand variable)	Independent (right-hand variable)	Coefficient Relating IndependentVariable to Dependent Variable	R^2
Figure 6.3	Fraser Institute overall rating in past 4 years	% of all students at Level 3 or 4 (both grades)	0.12 (0.003)	0.95
Figure 6.4	Fraser Institute overall rating in past 4 years	household income (1995 $ thousands)	0.0000840 (0.00000917)	0.59
Figure 6.5	average Z-score over Grade 3 and Grade 6	household income (1995 $ thousands)	0.00000147 (0.000005)	0.10
Figure 6.6	Fraser Institute overall rating in past 4 years	average years of parental education	1.11 (0.12)	0.58
Figure 6.7	average Z-score over Grade 3 and Grade 6	average years of parental education	0.26 (0.07)	0.19

Note: Shading indicates a coefficient (standard error) that is not statistically different from zero.

from the assessment written three years earlier. The estimated equation is:

$$\text{Grade 6 result} = b_0 + b_1 \text{ Grade 3 result 3 years earlier} + b_2 \text{ parental education} + e_5 . \qquad \text{(A6.5)}$$

The estimates of the coefficients b_1 and b_2 are positive. Two sample estimates of these equations presented in the Fraser report are found in Table A6.3. The R^2 in these equations are large: 0.5 or greater. The concept behind equation (A6.5) is that, if the school is a "success," its Grade 6 results will improve over the Grade 3 results by more than at other schools. A "good" school is identified when the value of e_5 is large. In Table A6.4, the letter A indicates that the value of e_5 falls in the first quartile of all values of e_5 in the sample of schools used to estimate the equation; B indicates e_5 falls in the second quartile; C in the third quartile; and D in the fourth quartile. Since

Table A6.3: ***Examples of Fraser Institute "Value-Added" Regressions***

Dependent (left-hand variable)	First Independent (right-hand variable) Coefficient (standard Error)	Second Independent (right-hand variable) Coefficient (standard Error)	R^2
Grade 6 reading score, 2002	Grade 3 reading score, 1999 0.37 (0.017)	average parental education 0.25) (0.014)	0.33
Grade 6 math score, 2002	Grade 3 math score, 1999 0.32 (0.018)	average parental education 0.22 (0.015)	0.23

Source: Cowley and Easton 2003, 253.

there are two Grade 6 assessments (reading and mathematics) in each of the 2000–01 and 2001–02 academic years used to estimate equation (A6.5), most schools have four letter grades. Although the concept sounds very similar to the Z-scores, there is no obvious pattern between the Z-scores in Table A6.4 and the letter grades for schools. Why might this be the case?

My answer is that equation (A6.5) is unusual. By linking the Grade 3 group three years earlier to the Grade 6 group in the current year, the intent of the equation is to take into account the school's socio-economic context as well as the effects of having a better (or worse) than average class through their effect on the Grade 3 assessment results. The effects that are specific to the class are handled in assessing the "same" group of students in both Grade 3 and Grade 6. The school community effects are handled because both the Grade 3 class and the Grade 6 class would normally share similar parents.

I have a couple of comments to make about equation (A6.5). First, if the results in the Grade 3 assessment control for socio-economic context, then there is no obvious additional role for the parental education variable. After all, if the previous Grade 3 class had the same parents, then parents should have no additional effect. Many people would expect the estimate of b_2 to be zero. It is not.

Table A6.4: *A Comparison of the Z-scores and Fraser Institute "Value-Added" Grades for 60 schools in the Waterloo Region District School Board*

School	Average Years of Parental Education	Average Z-score Grade 3 and Grade 6	Fraser Institute Value-Added Measure
Mary Johnston	15.5	0.06	BBBB
Laurelwood	14.5	-0.09	AA
Harold W. Wagner	15.0	-0.16	BAAA
Keatsway	15.6	-0.09	CADC
Lexington	14.1	0.68	AAAB
Clemens Mill	14.5	0.81	CCDC
Woodland Park	14.5	0.27	BBB
Empire	15.5	1.49	BBBB
Northlake Woods	14.4	-0.02	AAAA
Westvale	15.6	-0.35	BBBB
Blair Road	14.4	1.08	BBCA
Breslau	14.2	0.61	CCBB
Hespeler	13.9	0.79	ABAB
Sandowne	14.6	0.33	BBBB
Lincoln Heights	14.3	-0.56	BBCB
Tait Street	14.4	-0.04	DCDB
Franklin	14.2	0.07	CDBD
Meadowlane	13.9	-0.69	DDDD
Linwood	11.5	-0.90	AABB
Bridgeport	13.7	1.06	ABAA
Wellesley	13.9	0.12	ABAB
Silverheights	13.5	-0.54	DCCC
Glencairn	13.3	-0.59	CACB
Hillcrest	13.3	0.26	CBCC
N.A. MacEachern	14.3	0.49	AABB
Ryerson	14.0	1.39	DBBC
Country Hills	13.8	-0.02	CACA
Sheppard	13.6	0.64	BCBC
Elizabeth Ziegler	14.3	0.90	ABBB
Mackenzie King	14.3	-0.26	CBDC

Table A6.4 - *continued*

School	Average Years of Parental Education	Average Z-score Grade 3 and Grade 6	Fraser Institute Value-Added Measure
Westmount	14.7	0.25	AB
Centennial	13.4	-0.34	AABC
Southridge	13.9	0.82	BCCC
Preston	13.9	0.54	CDBC
Alpine	13.2	-0.62	DCDC
Trillium	12.7	-0.56	DACB
Elgin Street	12.6	-0.26	CDCD
Smithson	14.4	0.44	DCBB
Grand View	13.2	-0.54	DCDB
Forest Glen	13.8	-0.65	CBDC
Stewart Avenue	12.2	-0.03	BCAB
Alison Park	11.7	-0.18	BAAB
Central	12.5	-0.28	BBAA
Coronation	12.7	-1.10	CDBD
Avenue Road	12.2	0.01	AAAA
Crestview	13.6	-0.66	CDCD
Forest Hill	13.1	0.55	CCBB
Suddaby	13.9	0.23	CCDC
Parkway	12.5	0.38	DDDD
Cedarbrae	13.6	-0.25	DDDC
Manchester	11.5	-0.94	CCBC
Prueter	13.1	-0.11	DBDD
Howard Robertson	12.9	-0.82	DCDD
Winston Churchill	13.7	-0.76	BABA
J.F. Carmichael	12.7	0.23	CCCB
Rockway	11.7	0.04	ABBB
Wilson Avenue	12.3	-0.41	BDCD
A.R. Kaufman	12.8	-0.12	DBCB
King Edward	12.4	-0.95	BCCD
Queen Elizabeth	11.7	-1.22	DDDD

However, Table A6.1 provides a clue in understanding the positive estimate of b_2 in equation (A6.5). The variables in the table that represent the association of education with assessment results have a different effect on the Grade 3 than on the Grade 6 results. In particular, having a large number of adults in the school community with some university education (the equivalent of a high average value of the number of years of parental education) is associated much more strongly with higher Grade 6 results than with higher Grade 3 results. Thus, for schools with many highly educated parents, equation (A6.5) would predict, through a positive value of the coefficient b_2, higher Grade 6 results. The context variable here is playing the role of controlling for the impact of parental education on the increase in Grade 6 scores. Yet it is only because the association of education with results is different in Grade 3 and in Grade 6 that the parental education variable is a useful context variable in this equation.

A second issue with equation (A6.5) is that our previous results indicate that the turnover of households in a school community is associated with lower assessment results, very strongly in the case of the Grade 3 results and with a weaker, but still significant, effect in Grade 6. Thus, if the school is in a community with a lot of student turnover, then the Grade 3 results will be lower and so will the Grade 6 results, but not by as much. This would look like "improvement" relative to the Grade 3 results without the school's having taken any action, and the larger the turnover of students, the better the "improvement" would appear to be. This would generate a lack of correlation between the letter grades and the Z-scores because the "improvement" would be from a lower base. Further, with high student turnover, the idea that class effects are controlled in looking at results from previous years would be less useful. In Chapter 3, we learned that between 2.3 percent and 51.9 percent of school community households moved within the past year, with the average being 13.5 percent. Thus, in many schools, the composition of a specific grade could change substantially over a three-year period from Grade 3 to Grade 6. We have already seen that year-to-year fluctuations in class composition are important. Since the value of would be larger in a high-turnover school and randomly larger or smaller in a high-turnover school when there is more variation within the class, it is possible that the values of e_s actually contain very little information. All the fluctuations from Grade 3 to Grade 6 are random, and the lower the Grade 3 score associated with a high-turnover school, the more random the fluctuations would be.

The Fraser Institute's attempt at a "value-added" approach has one further important weakness. Because its scores use Method 2 data, the easiest route to school "success" would be simply to exempt virtually no students from the Grade 3 assessments, which would deliberately lower average scores in Grade 3, then to exempt as many low-performing students as possible from the Grade 6 assessments in order to raise average Grade 6 scores. As long as other schools did not follow the same strategy, the school would then appear to have improved. Of course, other schools would see the apparent value of following the same strategy, and the entire assessment process would be perverted. As discussed earlier, any actual value-added measure would require the analysis of individual student data, which are not currently available in Ontario.[13] In short, the Fraser Institute's "value-added" effort, though valiant, does not appear to be successful.

---

[13] For two examples of the analysis of individual assessment data, see Sanders and Horn (1994) and Webster and Mendro (1997).

7 Listening to Principals, Teachers, and Parents at "Good" Schools

In Chapter 6, I explained the process of identifying a "good" school — one that, over a number of years, outperforms schools with similar socio-economic characteristics on the standardized assessments conducted by Ontario's Education Quality and Accountability Office (EQAO). As a follow-up to that statistical process, I felt strongly that it was necessary to select and visit a number of such "good" schools and to interview principals, teachers, and parents to see if it was possible to identify what made those schools stand out and to share those insights more broadly. Of course, underperforming schools are equally interesting from an educational point of view, but I leave the examination of such schools to their boards.

Finding and Visiting "Good" Schools

In this chapter, I present data I collected from visits to 13 "good" schools in June 2003. I want to make it very clear from the outset that the schools I was privileged to visit were not chosen randomly, for three important reasons.

First, the school had to be "good" in the statistical sense that I established in Chapter 6 — in other words, a school in which a significantly higher percentage of all its students perform at Level 3 or Level 4 on the assessments than do the students of a school with similar socio-economic characteristics.[1] Table 7.1 presents data on the socio-economic characteristics of the 13 schools I visited.[2] Recall from Chapter 6 that a school with a Z-score of +1.29 or higher performed better than 90 percent of comparable schools on the assessments. Of the 13 schools I visited, Z-scores for both the Grade 3 and Grade 6 test results exceeded 1.29 10 out of 13 times. In one school, the Z-score for Grade 3 was 2.68 but only 0.36 for Grade 6; another school showed the reverse, with a Z-score for Grade 6 of 2.56 but only 0.63 for Grade 3. Lastly, one school in a very low income area displayed respectable Z-scores of 1.21 for Grade 3 and 1.18 for Grade 6, which prompted my interest in visiting it.

The second reason the 13 schools I visited were not chosen at random is that my sample was severely restricted by the refusal of most school boards I contacted to allow me to visit "good" schools under their jurisdiction.

I was very naive at the beginning of this part of the project. I assumed that boards and administrators would be delighted that a careful and convincing statistical analysis had placed the assessment results of their schools in an appropriate socio-economic context and identified outstanding schools in their board relative to similar schools in the province. I thought boards would want to present their success stories to the rest of the province. I was spectacularly wrong.

Any research project at a school must be approved through a reasonable process at the relevant school board. I thus began by writing to eight school boards in the immediate vicinity of my residence

---

[1] Students at Levels 3 and 4 "demonstrate and communicate extensive to in-depth knowledge, understanding and application of [reading, writing, and mathematics] skills" (EQAO 1999, 10).

[2] The timing of the school visits required the use of variables from the 1996 census. The income variables from the 2001 census were not released until July 2003, and I needed to complete the visits before schools closed for the summer in June 2003. The results are not, however, sensitive to the choice of census.

Table 7.1: *Socio-Economic Characteristics of 13 "Good" Schools Visited*

School	Average Household Income	Unemployment Rate, Adults with Children	Living in Detached Home	Adults with Some University	Adults without High School Diploma	Living in Canada < 5 Years
	(1995 $)	(%)	(%)	(%)	(%)	(%)
1	44,837	3.6	88.1	9.1	65.3	0.5
2	45,051	8.9	31.9	15.1	56.5	6.0
3	48,225	6.3	72.6	9.0	61.3	0.2
4	50,608	9.2	75.1	11.6	55.4	0.7
5	53,575	7.1	88.0	16.8	57.2	1.0
6	55,711	8.8	39.5	16.1	51.7	2.8
7	56,144	6.3	92.3	14.3	54.2	1.2
8	56,569	7.0	76.9	18.4	50.9	1.0
9	59,741	3.5	62.9	15.6	53.6	2.2
10	60,661	3.9	91.1	13.9	51.2	0.2
11	61,058	6.4	61.3	30.2	38.6	2.7
12	61,370	1.7	67.4	17.3	55.6	0.5
13	86,562	2.9	80.6	41.3	30.8	2.4
Provincial average	58,829	8.0	65.5	21.4	49.0	4.8
(standard deviation)	(16,190)	(4.5)	(23.0)	(10.5)	(9.6)	(6.5)

Note: Socio-economic variables are derived from the 1996 census.

in the Kitchener-Waterloo area. I wanted a mix of boards: rural, urban, suburban, separate (that is, Catholic), and public. My request made it very clear that the proposed visit would consume no resources at the board level and very few at the school level. In each case, I asked to visit two schools, and each visit would require about 30 minutes of staff time, 30 minutes of the principal's time, and 30 minutes of time at a school council meeting. I asked for no observation time in classrooms or contact with students.

Much to my surprise, six boards immediately refused my request, usually without giving any clear reason. Undaunted, I expanded my search for cooperative boards more generally in the area of southwestern Ontario, including boards in York Region and Simcoe

Table 7.1 - *continued*

Official Language Spoken at Home	Moved in Past Year	Moved in Past 5 Years	Single-Parent	Aboriginal Status	Total Enrollment
(%)	(%)	(%)	(%)	(%)	
99.6	6.9	21.6	10.4	0.55	181
88.6	21.4	51.6	31.5	0.97	158
97.0	14.2	14.1	24.4	1.72	164
96.1	16.3	17.6	22.6	1.00	153
95.2	12.4	31.1	21.9	0.46	364
88.5	12.8	36.8	19.3	0.44	391
97.7	8.9	32.3	11.9	1.00	134
94.0	13.1	36.0	23.9	0.73	413
95.7	10.1	29.1	19.4	0.44	309
97.6	8.6	24.4	13.0	0.02	246
95.9	18.2	45.3	28.2	0.25	154
92.3	10.6	31.2	14.7	0.33	197
93.1	11.2	17.3	12.6	0.04	457
91.0	14.4	42.5	21.3	1.5	331
(10.7)	(4.9)	(10.1)	(8.8)	(1.5)	(167)

County.[3] Eventually, the following 13 schools in 6 boards (in no particular order) agreed to participate in my survey:

- Mildmay-Carrick Central School, Bluewater District School Board;
- St. Teresa (Elmira) School and St. Aloysius School, Waterloo Region Catholic District School Board;
- West Lynn School and Boston-Townsend School, Grand Erie District School Board;

---

[3] I chose to avoid school boards in the Toronto area since I did not want to become embroiled in the complicated politics of these boards as they were in spring 2003. I am also aware that the Toronto District School Board has some of its own research capabilities in this area.

- St. John the Baptist School, St. Luke School, and St. Martin of Tours School, Hamilton-Wentworth Catholic District School Board;
- Northdale Public School, Leesboro Central School, and Jack Chambers School, Thames Valley District School Board; and
- Parnall School and McKay School, Niagara Regional District School Board.

I am very grateful for the permission of these boards to visit their schools and for the co-operation of staff and parents at these schools. All visits but one were a delight, and even the one less than cordial visit (for reasons I describe below) certainly made for a lively and interesting day. I learned a lot from the school visits.

The third reason the schools I visited were not randomly chosen is that, even at boards where I was welcome, I decided not to visit *all* the available "good" schools. Rather, I wanted to visit schools with a range of characteristics — those serving communities whose socio-economic factors predicted the school would have a lower score as well as those whose score was predicted to be high. Moreover, in two cases, I obtained permission from boards to visit particular schools but the schools themselves denied my request.

In summary, I must admit that the research design related to the choice of schools to visit was far from ideal — my visits were short and to a small sample not selected in a random process. Future researchers should strive to use other methods. In a better world, a random sample would be drawn from all "good" schools in Grade 3 and all "good" schools in Grade 6. In an even more ideal situation, many observations of behaviour and activity over long periods of time would be made at a random sample of all types of schools and then correlated to schools' Z-scores. Nevertheless, the visits I made were those that were feasible given existing constraints on time, resources, and permissions. One has to start somewhere.[4]

---

[4] A substantial economics literature links school resources to school outcomes, although, in most cases, the link is weak. When data on individual students and teachers are available, some studies have found that individual teachers can have a significant effect on learning outcomes. The economics literature,...

A Typical School Visit

A "perfect" school visit would have consisted of meetings with the principal, all staff, any "key" staff member identified by the principal (if the principal desired and the staff member agreed to be interviewed), and, finally, a selection of parents at a school council. My 13 visits, however, were not all "perfect." In particular, it proved complicated to meet parents. With more timely provision of EQAO assessment data and faster permissions from boards, I might have managed more visits with parents. In the event, however, I was obliged to confine my visits to a three-week period in June 2003, and during that time some schools did not hold school council meetings. I did manage to meet with seven full school councils and with a number of other parents whom principals had arranged for me to see.

Other complications also prevented "perfect" meetings in some cases. At one school, either the administrators did not quite understand or I did not communicate clearly that I wanted to meet with *all* staff, not just the Grade 3 and Grade 6 teachers. Accordingly, my meetings at that school consisted of lengthy interviews with just Grade 3 and Grade 6 staff as well as with the Learning Resource Teacher (LRT); as will be apparent below, this is almost certainly the wrong approach. At another school, a work-to-rule labour action prevented a scheduled meeting with staff as a whole, although, over the course of the day, I was able to meet the majority of the staff as their schedules permitted. At some schools, I was unable to meet with a "key" staff member identified by the principal; indeed, for reasons that become clear later, the concept of a "key" staff member

---

Note 4 - cont'd.

...however, focuses on measurable inputs to the education process, such as dollars spent per school or the number of teachers per student; the literature does not typically identify school activities that are correlated to higher or lower scores. For a very good summary of the economics literature, see Hanushek (2002). For a provocative discussion of the statistical issues involved in identifying "best practices," see Hanushek (2004).

turned out to be less useful than I had expected. In several schools, the principal had been in the job for less than a year and was perhaps less able to offer useful information about the school and its practices than a more experienced administrator — although in each case, I was able to interview the former principal.

Nearly everyone did their best to help me and to answer my questions. I found that serving homemade cookies or muffins in the staff room was a successful strategy for meeting teachers and getting responses. I did my best to listen. I promised teachers and administrators that, in this chapter, I would accurately and dispassionately record what they had chosen to tell me — I hope I have succeeded. As one teacher said to me: "You could have hardly chosen a more controversial topic of research than EQAO testing."

What Principals Told Me

As part of my visits to the 13 schools, I interviewed 17 principals[5] using the questionnaire found in Box 7.1. Some chose to fill out the form, others had me record their verbal responses. In each case, principals were asked to identify specific activities that took place at their school that contributed to EQAO success; any differences between the staff at their school and staff at other schools in which they had worked; and any other differences, not associated with staff, between their school and other schools in which they had worked. The questionnaire also asked about the school's goals and how they were being implemented; however, the responses to these questions were much more difficult to interpret and I made no use of them.

Of the 17 principals, 13 identified the "team" aspect of their teaching staff as a very important factor in the school's success. I heard the word "team" many times in the course of these visits. It usually meant the group of teachers that worked in the primary

---

[5] Two were actually vice-principals, one of whom was filling in for a principal on a school trip, and three were former principals. All the principals involved were given an opportunity to comment on a draft of this chapter; none offered corrections.

Box 7.1: ***Questionnaire for Principal***

About yourself:

How many years have you been a principal?
How many years have you been at this school?
How many years did you teach? What grades?

An initial question: Your school performs better than similar schools in the Grade 3 and Grade 6 tests administered by the EQAO.
Did you know this before my research visit? (circle) YES NO

Please try to list three activities that occur at this school that you believe help increase your EQAO test scores.

We all recognize these activities take some time and resources. Can you identify activities that are *not* emphasized at your school as a result of choosing to spend time and resources as described above?

About your school:

About your staff:

Can you identify differences between the staff at this school and the staff at other schools you have been involved with that contribute to the success of your school in EQAO testing? Please try to list up to three such differences.

How, in your view, did these differences in staff come about?

Are there key staff persons who contribute to your success in EQAO testing?

Describe their characteristics.

What specific activities do they undertake at your school?

Would you be able to identify this person or these persons and arrange for me to speak to them for 15 minutes sometime later today?

About your school community:

Please think back to other schools you have been involved with. Can you identify differences in this school community that lead to your success on EQAO testing? Please try to list up to three such differences and explain why they are important.

Please describe the three most important goals you set in your operation of this school community?

Give three examples of ways you act to implement those goals.

division (junior kindergarten to Grade Three) and the group that worked in the junior division (Grades Four to Six). As we will see, the functioning of these two groups of teachers as teams is the factor teachers themselves also identified as the most common difference between a successful school and other schools.

From this evidence, I conclude that teachers working as a team by division is the crucial factor in school success on assessments. Principals viewed the teams as being able to identify, over the sequence of grades, students who needed specific skills that would allow them to improve from Level 2 to Level 3. In addition, an important role frequently was played on teams by an implicit teacher-leader — an experienced mentor or partnering teacher or perhaps a teacher with EQAO marking experience. Teams were also able to identify for the principal the specific teaching resources needed to prepare for the assessments. The most commonly mentioned resources were "primary mathematics manipulatives" (hand-held objects that help students understand abstract mathematics concepts) and "levelled books." Indeed, the latter were mentioned so often that I began to understand their role clearly. These are sets of books in the primary division where up to 30 levels of reading skills are identified. As students proceed through these levels, they are supposed always to be reading at one level above their capabilities. The analogy is finding a tennis or squash partner who is just a little bit better than you — if your partner is a lot better, then the game is so one-sided you do not improve; if you are a lot better than your partner, your victory comes too easily and you do not improve. Analogously, students who read a book that is just a bit "too hard" for them have the opportunity to improve their reading skills.

I then asked principals how such competent teams came into being at their school, but only one was able to answer this question. When that school opened just 11 years ago, more than 200 teachers had applied for the 20 available jobs — in the phrase of one teacher, they were "handpicked" by the school's first principal precisely because they were interested in the elementary education process as a team. During the four years of assessments studied, this "handpicked" staff constituted the numerical majority of staff at that school.

No other principals offered any clear insights into how such staff "teams" could be assembled. Normally, new teachers appear to arrive at a school without principals' playing a large role in their selection process. Many teachers choose to stay at a school for long periods, particularly when the team is successful. In rural schools, teachers often stay for a very long time because switching schools may involve a long-distance commute. Principals were, however, able to identify some specific activities that they believe contribute to team building at their school.

Some principals identified the importance of allowing teachers specific blocks of time to develop co-ordination within the primary and junior divisions. In some schools, teachers' time was freed up by principals' covering the relevant classes themselves. In other schools, primary or junior division teachers worked closely with their principals in ordering teaching resource material. Five principals talked about how they had managed changes in the evolution of the EQAO assessment process over the years, citing the usefulness of books on leadership through change and their respect for their teachers' skills and knowledge. These principals also talked about how they persuaded their staff to recognize the importance of the assessment process, and how they made extensive use of the resources at the board level, sending their staff out for in-service training as frequently as possible. At two schools, the LRT played a huge role in staff development.

As for the activities of their students, virtually all the principals I interviewed said that student preparation for the assessments was key. These principals usually had a clear plan to "get ready for the test," although they did not consider this to be "teaching to the test" — indeed, many made it very clear they did not agree with the use of that term. Test preparation included practising answering multiple-choice questions (particularly in Grade 3, where students are not used to this form of testing), learning the icons on the EQAO assessments, and using previous tests for practice. In more recent assessments, teachers have had more example questions and rubrics to explain how tests are marked. As one principal put it, "test wisdom" involves learning these marking schemes and teaching them to students.

All of these activities build students' confidence as they approach the assessment. Only one of the 13 schools treated the assessment in a manner remotely similar to that of a high school or university examination — that is, as a situation in which students would perceive considerable pressure related to the outcome of the assessment. Instead, schools viewed the assessment as a low-key activity in which well-prepared and relaxed students would do their best. One principal told me of a Grade 3 teacher who taught her class to giggle at the letters EQAO.

For 10 of 17 principals, early intervention and assessment in the primary reading process was a key activity. Weak readers in all three primary grades were given extra help and support, usually with the help of volunteers often drawn from among parents. In two schools, the LRT played a particularly important role in organizing and training volunteers. Another school restricted volunteers, beyond kindergarten, to working in classrooms in which their own children were not present, in the belief that children were distracted by having their own parents in the room and that it was important for volunteers to understand they were helping the school community, not just their own children.

Several principals at schools with lower socio-economic characteristics displayed particular creativity in obtaining and using volunteers from their community by going beyond parents to make effective use of students from nearby high schools, co-op placements from high schools, Early Childhood Education students from local colleges, and university students and retired teachers. Crucially, such volunteers were given a clearly identified role in the early intervention process, as well as training and guidance.

Other activities principals identified as important to success on the assessments were much less systematic. Principals at six schools mentioned the importance of good extra-curricular activities, such as musical activities, athletics, or field trips. Three principals mentioned the importance of regular recognition of achievement, both academic and non-academic. Four principals emphasized the importance of "straight" classes, while another felt that "split" classes were an advantage. (In a straight class, all students are in the same numerical grade;

a split class contains students from more than one grade.) At three schools, principals identified the importance of their own skill at using the maximum "accommodations" allowed under the EQAO assessment process. (An accommodation means that, although a student taking the reading test cannot be given any assistance, in the mathematics test it is permissible to read the question *verbatim* to a weak reader.) Several principals felt not all schools took full advantage of such accommodations. Five principals described the importance in their schools of identifying weaknesses from previous EQAO assessments and then incorporating appropriate corrections into the teaching process. One principal, however, rejected such a process as "too intense."

My final comment about what principals told me relates to those who work at schools in communities whose lower socio-economic characteristics predict lower assessment scores. Two such principals identified as their most important task the creation of a school environment in which learning could take place. As administrators, they said that they expended a great deal of energy and resources meeting the broad social and discipline needs of their school community, and that their teachers were able to function better when these needs were met effectively. Teachers and other co-workers greatly appreciated the role of these administrators in creating or improving the learning atmosphere, which was identified as key to the success of these lower-income schools. There is no question in my mind that principals at "good" lower-income schools are strong leaders.

What Teachers Told Me

As with principals, teachers at the 13 schools I visited were asked to respond to a questionnaire (see Box 7.2) that asked five key questions. First, teachers were asked to summarize, in one sentence, why they felt their school did so well on EQAO assessments relative to similar schools. Second, teachers were asked to identify three activities undertaken by staff or students that they felt contributed to higher achievement on assessments. Third, teachers were invited

Box 7.2: *Questionnaire for All Staff Members*

About yourself:

How many years have you been a teacher?
How many years have you been at this school?
What grade do you teach at this school?

About your school:

My research shows that your school performs better than similar schools on the EQAO testing. Does this surprise you? (circle) YES NO

Can you provide a one-sentence answer to the question:
What is the most important reason your school does better than similar schools elsewhere in the province on EQAO assessments?

Name three activities that occur at this school that you believe help increase your EQAO test scores.

These activities take some time and resources. List up to three activities that you observe are *not* emphasized at your school as a result of choosing these activities?

About your school community:

Please think back to another school you have been involved with. How long and in what way were you involved with that community?

Can you identify differences that contribute to achieving learning outcomes between the staff and leadership at this school and the staff and leadership at the school you identified above?

Can you identify any other differences between the previous school community identified above and your current school community that lead to success in the current community? Please try to list up to three such differences and explain why they are important.

Are there other comments you would like to make concerning the success of your school?

to consider what other activities the school might not have undertaken because time and resources had been diverted to preparing for assessments. Fourth, teachers were asked to list differences between their current (relatively successful) school and other schools in which they had taught or been active. Finally, all teachers were given the opportunity to make any other comments they wished.

A questionnaire given to "key" staff, as identified by principals, turned out not to be very helpful (see Box 7.3). Both principals and teachers consistently responded that strong teamwork within the primary and junior divisions, rather than the presence of particular staff members, was the major factor in generating strong assessment results.

Teachers' Observations about Their Own School

Table 7.2 summarizes the responses of 136 teachers to the question: What is the most important reason your school does better than similar schools elsewhere in the province on EQAO assessments? Many teachers gave more than one reason, frequently responding with long, compound sentences and even point-form lists.

Among all responses, the reason most often cited was the presence of excellent teachers. The second most often cited reason was thorough test preparation, or "test wisdom," closely followed by the importance of strong teamwork among teachers in both the primary and junior divisions. Other frequently cited reasons included high expectations of students, good professional development and training, and strong parental support. Twelve teachers felt that the fact their school was in a high-income community was the most important factor in achieving good results — despite my having clearly explained to them that their school was being compared with schools with similar socio-economic characteristics.

Table 7.3 summarizes the responses teachers gave when asked to identify three activities, undertaken by staff or students, they felt contributed to higher test scores. It is quite clear that these successful schools devote considerable time and resources to test preparation: students practise questions, solve math problems, undertake a lot of math drill, and are taught the necessary test vocabulary and icons. Most of these "good" schools also have an active early literacy program: 28 teachers reported the importance of reading-reward programs, 11 stressed group responses to reading, while a similar number cited two particular literacy programs, the "Four Blocks" and "Jolly Phonics," as helpful (these programs had been adopted in four

Table 7.2: *Why Teachers Thought Their School Performed Well*

Teachers' most frequent one-sentence responses to the question:

"What is the most important reason your school does better than similar schools elsewhere in the province on EQAO assessments?"

Reason	Number of Responses Containing This Reason
Excellent teachers	40
Good preparation for assessment	31
Teamwork within divisions	28
High expectations from students	21
Good training in new curriculum	17
Strong parental support	16
A high-income community	12
Lots of good resources	7
Strong leadership from principal	7
A small school	5
Lots of homework	5
Strong leadership from other teachers	4
Continuity in staff	3
French immersion at school	3
Communication with parents	2
Direct response to EQAO results	2
Many exemptions granted	1

Note: Some teachers provided more than one reason.

of the six boards I visited since the EQAO assessment process began). Some teachers also cited the strength of extra-curricular programs (music, drama, and athletics) as critical to the success of their school, and a large group of teachers identified daily or weekly communication with parents about classwork as crucial.

As far as their own activities are concerned, it is not surprising that many teachers mentioned planning and co-ordination activities within divisions as crucial to success on assessments. A significant number also cited the importance of their school's response to weaknesses identified in previous EQAO results. Teachers also found professional development in the area of literacy to be particularly valuable, with several mentioning in particular the excellent literacy training offered by their board, of which their school had taken full advantage. Finally, nine teachers mentioned the abundance of learning resources.

Table 7.3: *Activities Teachers Identified as Increasing EQAO Assessment Scores*

Activity	Number of Mentions
Students	
Related to test preparation	
Direct practice	19
Math problem solving	20
Vocabulary related to assessments	10
Math drill	9
Math manipulatives	5
Related to literacy	
Reading reward program	28
Shared responses to reading	11
"Four Blocks" and "Jolly Phonics"	11
Early intervention on literacy	9
Reading buddies (upper grades or parents)	7
Other activities	
Extracurricular activities	16
Direct, clear, regular (daily or weekly) communication with parents	19
Staff	
Within division team planning	40
Direct responses to weaknesses identified in previous EQAO assessments	16
Good professional development	12
Using good teaching resources (levelled books and math manipulatives)	9

Note: Responses summarized from 136 teachers at 13 schools.

As an economist, I felt I had to ask teachers whether there were other useful activities that the school had *not* undertaken because time and resources had been diverted to preparing for tests. I found that I nearly always had to explain the question by way of an example. I would say to the group: "Suppose my class does so much finger painting, it does not have time to read. Then I have given up reading activities in favour of finger painting." Table 7.4 summarizes teachers' responses. The most common answer, from 57 of 136 teachers, was that the school had given up nothing — that none of the activities they had undertaken to improve assessment scores had meant a reduction in other activities that would otherwise

Table 7.4: ***Activities Teachers Identified as Not Empasized in Favour of Improving EQAO Assessment Scores***

Teachers' most frequent responses to the question:

"The activities that increase assessment scores take some time and resources. List up to three activities that you observe are not emphasized at your school as a result of choosing assessment-increasing."

Activity Not Emphasized	**Number of Responses Containing This Reason**
None	37
Art	25
Fun and seasonal activities	15
Science	11
Teacher's personal time	8
Music	8
Primary French	5
Physical education	5

Note: Responses summarized from 136 teachers at 13 schools.

have taken place. Many other teachers, however, identified losses in art, fun and seasonal activities, science, music, personal time, and physical education. One board had made a decision to begin elementary French in its schools only after Grade 3, which some teachers perceived as having allowed more time for English literacy in the primary division. A number of teachers also volunteered the opinion that, during the period of approximately two weeks to a month just before the assessment, non-literacy and non-numeracy material needed for the assessment crowded other material out of the curriculum. I return to this issue below.

Teachers' Comparisons of Schools

My questionnaire also asked teachers to compare their current school with other schools in which they had worked (without identifying them). To my surprise, the proportion of teachers that had no significant experience in other schools was quite large. About 20 percent had been teaching for only a short period and, except for practice teaching, all of their experience had been at their current school. Another 20 percent or so had taught at their current school for more than ten years — some for more than 20 years — and felt they could not, therefore, compare schools. Of the other 60 percent of teachers, however, I asked for comparisons in two areas: staff and leadership, and anything that was not staff and leadership. The responses are quite revealing.

Teachers' comments on differences they identified in staff and leadership mirror the importance of the now-familiar theme of teamwork within primary and junior division. For example: "I had no idea what other teachers were doing at my other school; we all just shut our doors"; "there was no teamwork where I was before"; "at the other school, there was division among the staff, not much communication, and lots of discipline problems."

The last comment leads directly to teachers' views of the role of administrative leadership. So far, in teachers' responses to questions about their school, the role of principals had not figured prominently — indeed, even principals themselves did not stress their role. However, when it came to factors that distinguished their current, successful school from a previous, less successful school, more than two dozen teachers identified the principal as a key difference. Their comments included: "[the need for] consistent discipline"; "amazing administration with very high expectations"; "at the other school, there was no manager"; "the experience with this principal was positive and when I had the opportunity to work with him/her again, I did"; "my other school did not have appropriate leadership"; "the principal demonstrates leadership."

Teachers also appreciated a principal who provides resources, training, and support: "principal provides the needed resources"; "principal praises teacher effort outside of teacher appraisals." When asked what it was that principals actually did to increase assessment results, however, teachers had some difficulty responding. They had no trouble identifying the key characteristics of a good principal, however, and knew when the leadership had improved from one school to the next. In two schools, teachers identified the LRT as a key leadership person in terms of teacher training and mentoring.

Many teachers, however, seemed not to understand that I had picked their particular school because it had performed better on the assessments than *similar schools*, and continued to suggest that their school had done better because its community had better-educated, higher-income parents, fewer ESL students, more stable families, and the like. Perhaps my line of inquiry was insufficiently clear in this respect, since all previous discussion clearly had been about similar schools. Some teachers thought that their school had an advantage because of its small size.

Bits and Pieces from "Key" Teachers

In asking principals to identify "key" teachers, my intent was to interview teachers who had made a strong contribution to a particular activity. My questionnaire asked these teachers for details about activities that had contributed to success and about their role in these activities (see Box 7.3).

It became apparent, however, that this part of the interview process was not well designed. I had not made it clear enough that a "key" teacher did not have to be one who taught a grade that took the assessments, with the result that I was frequently sent to see a Grade 3 or Grade 6 teacher. When asked if he or she was surprised to be singled out, the "key" Grade 3 teacher typically responded: "I am surprised to be singled out.... Everything the children have learned has been the result of what they did in junior kindergarten, kindergarten, Grade 1, and Grade 2" or "It is not just the Grade 3 teacher that prepares the children for testing. Working as a team is key." In fact, in the primary division, there are no "key" teachers, and principals and teachers themselves typically give full credit to the division as a whole, working as a team. The only modest exception might be the presence of a teacher with experience to share of marking EQAO assessments.

In contrast, several schools identified the Grade 6 teacher as "key"; moreover, many of these teachers were inclined to agree that their efforts as individuals, rather than as a team, had made a

Box 7.3: *Questionnaire for Staff Member(s) Identified by Principal as Key to EQAO Success*

As you can tell from the heading on this questionnaire, your principal has identified you as particularly important in the success that your school community has had in the EQAO testing process for Grade 3 and Grade 6 students. I really appreciate your time spent with me now.

About yourself:

How long have you been a teacher in total?
How long have you been at this school?
What grade do you teach?

Are you surprised that your principal has singled out your contribution?
(circle) YES NO Explain why you circled your choice.

Name three activities that you believe helped to increase test scores. In each case, please tell us the activity, why you believe it contributes to the success of the students at this school, and how you helped to implement that activity.

These activities take some time and resources. What activities do you feel are not emphasized at your school as a result of choosing these activities?

If you had more resources, what is the next activity that you would propose to improve learning outcomes? Why did you choose that activity?

You could implement this new activity if you gave up another activity. What activity would you reduce in order to implement your suggestion?

difference to their students' Grade 6 assessment results. Indeed, it is my own experience that an adult leader can have a profound influence on adolescents in this age group, and I expect the strong Grade 6 teachers know that as well. They used words about themselves such as "coach," "mentor," or "model"; assessments and results were "our opportunity to shine" and "our scores."

Asked to name mathematics activities they found useful, these Grade 6 teachers reported: timed activities, practising math tests, constructing tests based on large-scale investigations and spending time to correct them, math investigations related to the exemplars, and a math program called Quest 2000. In reading and writing, they cited, among other things: writing biweekly essays, using the book *Preparing for the Grade 6 Provincial Assessment — My Ontario Reading Workbook*, using sample assessments and explaining how to improve answers, familiarizing students with the style of questions on the assessments, setting timed writing activities, and having students practise correcting sentences on the chalkboard.

Some Grade 6 teachers also communicate effectively with parents about their expectations of students in their class: "I compose and distribute the Friday letter every week (consistency is key) and track that it has been received by each family" and "Every day at the end of the day students copy their homework into their planner. This planner is taken home, viewed by parents, signed by parents, and returned the next day."

Teachers' Comments on the EQAO Process

In addition to specific questions, teachers were invited to make any further comments they wished on the EQAO process as a whole. Most chose to express themselves in the other parts of the questionnaire, but a few offered additional criticisms of the process.

For example, some teachers felt the assessments placed great pressure on students, and that testing with time constraints (especially in Grade 3) was an "unnatural situation" for these students A huge majority of teachers, both in writing and verbally, wanted the assessment period moved to later in the school year — to late May or even the beginning of June. Many felt that to get the math and literacy material done by April created a rush.

Some teachers felt that students did not receive the results of the Grade 6 assessments quickly enough, and that, for students, long-term feedback on the assessments was no substitute for short-term feedback from the classroom teacher.

One School's Negative Response to My Visit

My visit to one school went poorly. As at other schools, the teachers came to a meeting scheduled over a lunch hour and I explained the project to them and the reason for my visit. However, several voiced strong objections to completing my questionnaire or to discussing their school's good results on the assessments. They told me they wanted no part of a process that could be perceived in any way as supportive of the EQAO assessments. Some were extremely suspicious of my motives for undertaking the study and did not want to be associated with a project that was to be published by the C.D. Howe Institute.[6] Despite my assurances that public policy research was simply a part of my university job and that I had not been hired by the Ministry of Education, the school board, or the teachers' federation, the vast majority of staff refused to complete my questionnaire. Since teachers in other schools had been willing to share their insights, while recognizing that the assessment process is not perfect, I was taken aback by this response.

With this one school, therefore, I felt I had to adopt a different strategy. To this school's teachers, I subsequently posed two written questions that, perhaps unfairly, I did not put to the teachers of any other school. First, I asked them to list up to ten negative consequences for staff or students from the EQAO process. Their answers included:

- "too long, too disruptive"; "five days of testing is too long and there is too much pressure for children"; "there is another board-wide literacy assessment as well; this is too much assessment";
- "the failure of parents to realize this assessment is just one week of the year, not how their child does throughout the rest of the year";
- "multiple-choice questions require particular preparation, which is not age-appropriate in Grade 3";
- "students must spend too much time becoming familiar with a very artificial and limited EQAO vocabulary"; "there is an inordinate amount of time spent on test preparation activities"; "the curriculum becomes too rigid and there is time-pressure to complete the curriculum for the assessment"; "there is not much

---

[6] There was significant confusion in the group between the C.D. Howe Institute and the Fraser Institute, confusion I could not dispel.

opportunity to run with a teachable moment"; "this is not the way I want to teach";

- "subjects like Science, Social Studies, Music, French, and Physical Education lose out to prepare for testing";
- "disadvantaged students arriving at a school might be viewed by teachers and administrators as students who would lower test scores and thus would not be welcome";
- "these tests provide ample opportunity for administrators to become obsessed with data; they lose sight of the human beings in the classroom";
- "teachers already know that the biggest factor in determining whether a child does well on the EQAO assessment is the socio-economic background of the child"; "low-scoring schools get a bad reputation and become the domain of younger, inexperienced teachers who take the jobs they can get"; "excellent teachers may choose not to go to low-ranking schools."

Some teachers answered this question not by listing negative consequences of EQAO assessments but by offering criticisms of the assessment process itself, including: inconsistent administration and marking of the assessment; the fact that parents look at EQAO results in choosing a school or an area in which to live; the small sample size in a given year of results (so that one student represents a 5 point shift in the percentage at Level 3 or Level 4 in a 20-student class); and year-to-year variance in the tests themselves.

In my second question for the teachers at this school, I wanted to solicit their views about the cost of the EQAO process. I asked them to assume that each assessment costs $200 per student and that, in a school with equal numbers of students in each grade from junior kindergarten through Grade 6, cancelling an assessment would free up an extra $50 for each student — for a class of 25, this would be an extra $1,250 per year.[7] I then asked teachers to list between five and ten better ways to spend these assumed savings.

---

[7] In fact, this assumption, really just a guess made at the time, proved to be an exaggeration. Later research, which I presented in Chapter 2, revealed that the actual cost per elementary assessment is between $50 and $100.

Some teachers responded by listing sensible items that such relatively small sums could reasonably purchase: new (levelled) books, a new computer in the classroom, physical education equipment, math manipulatives, workbooks, spending on after-school programs to help disadvantaged children learn, more field trips, manuals for teachers, new audio-visual machines, new textbooks. However, many teachers seemed to believe, quite unrealistically, that $1,250 per class per year somehow could cover the cost of a full-time librarian, a teacher-librarian, more teachers generally, a music teacher, new desks for all students, a new gym, nine days of professional development instead of four days, or more educational assistants.

Indeed, responses both from teachers at this school and those at other schools led me to conclude that EQAO has not done a good job of communicating to teachers the actual costs of the assessment process — for example, no such information is posted on its web site.

What Parents Told Me

At seven of the thirteen schools I visited, I attended formal meetings of school councils; I also spoke to parents on the school councils of two other schools. I should note that no council meetings were scheduled during the time I visited five other "good" schools. Moreover, I cannot claim that the parents I met were random representatives of parents at their school; rather, they were motivated parents who are interested in what is going on at their children's school, and their contributions to my survey no doubt are skewed accordingly.

As I had for the other groups, I prepared a questionnaire for parents (see Box 7.4). Once again, I asked them to identify school activities they believed increased assessment results. Their lists, while shorter, reflected those of teachers and principals. Parents in school communities with lower socio-economic characteristics, fully appreciated the importance of consistent discipline and expectations for students' behaviour and performance, an issue that simply did not arise among at higher-income schools. Parents at all schools valued regular, clear communication from teachers about homework

Box 7.3: ***Questionnaire for Parents***

About yourself:

How many years have you been sending children to this school?

In what grades do you currently have a child?

Are you a regular participant at school council? (A regular participant attends at least every second meeting) (circle) YES NO

Your school performs better than similar schools on the EQAO testing. Were you aware of your school's strong performance? (circle) YES NO

I want to ask you about activities you see in your school community that you believe contribute to high EQAO test results. These activities take some time and resources. Can you name an activity that you feel is *not* emphasized at your school as a result of your list above?

About your school community:

Please think back to another school or schools you have been involved with. How long and in what way were you involved with that school community?

Can you identify differences between the school community identified above and your current school community that lead to success in the current community? Please try to list up to three such differences and explain you believe they are important.

About the goals of your school community:

Please describe your beliefs on up to three goals of your school community and give examples of ways in which you see these goals carried out.

and other expectations. They were fully aware of the thorough preparation for the assessment process that teachers undertake, and they appreciated their children's ability to approach the assessments in a comfortable, relaxed, and confident way. Some parents mentioned test anxiety, but the issue did not dominate discussion. Parents were also aware of early literacy programs at various schools and thought early intervention for weak readers was important. They also valued reward programs in reading and spelling, and parents at one school cited the importance of a "math club" for those needing special help in that subject. Parents at several schools mentioned the use of trained volunteers in early literacy programs.

In their responses, parents often took the opportunity to say what they liked about their school. They stressed repeatedly their view of the importance of extracurricular activities — music, athletics, trips, chess clubs, Scrabble clubs, dramas, and the like — in building a strong school community and in meeting their children's needs in a variety of dimensions. At the five Catholic schools I visited, parents cited the importance of the role of their school as a faith community for their children and appreciated principals who supported that role for the school.

Few parents were able to compare their school with others since most had not been involved in councils at other schools. The few who answered this question mentioned that staff at their current school worked better together, that the principal was more supportive of staff, and that the current school made better use of volunteers, communicated more clearly and more effectively with parents, and had better strategies for dealing with behaviour issues.

I asked parents how they perceived the goals of their school were being set and met, but this fairly difficult question did not generate many useful responses. Some parents observed that responding to specific weaknesses that EQAO had identified was a goal of their school. Others felt that the key goal of their school was to build a real partnership with parents. Parents in school communities with lower socio-economic characteristics indicated that a key goal of their school was to provide an environment where learning could take place.

As with staff, few parents saw a significant cost if their schools chose to divert resources to achieve better assessment results. Some, however, noted the tradeoff their boards had made between increasing time spent on primary literacy in English and starting elementary French in Grade 4. Several parents at Catholic schools felt that the increased time spent on literacy had meant a reduction in time spent on religious instruction. A very few parents regretted that the music component of education had been reduced.

Finally, parents had only scattered complaints about the EQAO process itself. One group said it had tried to grade student writing and had found it difficult to do; therefore, it had little confidence in the accuracy of the EQAO grading process and felt that individual test results could be misleading. Some parents had heard teachers' complaints that preparation for the assessments had taken time away from other activities, but they were unable to identify those activities. A very small minority felt their school put "too much emphasis on the test."

Summing Up the School Visits

I would like to reiterate that, in this chapter, I have tried very hard simply to report what principals, teachers, and parents told me during my visits to their schools. Importantly, those I heard from were not a random sample of participants in the Ontario elementary school system. Rather, my sample was limited to participants in a few "good" schools that did much better than similar schools on the same assessments over a four-year period. In a sense, these schools selected themselves by their willingness to support my research and allow me to visit them. In all cases but one, administrators, teachers, and parents were willing to share with me their views on why their school performed so well. I appreciate their thoughts, and I hope the reader found them to be as interesting as I did.

Chapter Summary

Chapter 6 discussed a methodology for identifying a "good" school — one that outperforms schools in communities with similar socio-economic characteristics on the EQAO assessments over a number of years. This chapter recounts my visits to 13 such "good" schools in half a dozen boards in June 2003. As part of each visit, I distributed questionnaires to administrators, teachers, and parents asking them to comment on the activities the school was undertaking to raise assessment results, and to identify differences, if any, between this successful school and any other school of which they had significant experience.

Principals and teachers told me that the key feature of a successful school was the ability of its teachers in each division to work together as a team. As a team, teachers can more readily identify students who need to add particular skills to their repertoire in order to achieve better results on the assessments. I am forced to conclude, therefore, that this feature is not common to all Ontario elementary schools.

Teachers identified a number of specific activities as particularly helpful. For the Grade 3 assessment, these included specific test practice and preparation, levelled books, mathematics manipulatives, and various literacy programs. For the Grade 6 assessment, teachers cited thorough test preparation, including timed writing and mathematics problem solving, the clear communication of homework expectations to parents, and, in some cases, strong leadership from individual Grade 6 teachers.

A "good" school makes effective use of volunteer resources. A number of schools made an effort to train volunteers. One school excluded volunteers from classrooms in which their child was a student. At several low-income schools that performed very well, principals sought out and obtained volunteer and other resources from the local community. At the successful schools that were predicted to have lower assessment results but did not, participants acknowledged the clear role of their principal in creating a disciplined and supportive atmosphere for learning.

At one school I visited, teaching staff refused to participate in the questionnaire. In its place, I asked them to tell me what they saw as the negative consequences of EQAO assessments and how they would rather allocate the money now spent on the tests. Answers to the first question revealed the difficulty some teachers had in separating problems they themselves may have had with a particular test and their views of the assessment process as a whole. Their views also underscore the need to interpret assessment results in the school community's socio-economic context. Answers to the second question highlighted a common, powerful myth that greatly exaggerates the actual magnitude of resources used in the assessment process. Seeking the views of teachers more generally on these issues would be helpful.

The chapter's main message is that it is possible to identify "good" schools and visit them to find out how they do it. The answers of my small sample of participants in Ontario's elementary school system are thought provoking and useful. Much more work of this type needs to be done.

8

Conclusions

In writing the last chapter of this book, I had two goals in mind. One was to present the traditional summary usually found at the end of a piece of academic writing. This is appropriate since the book is, first, an academic and statistical study of the results of the standardized assessments conducted by the Education Quality and Accountability Office (EQAO) in Ontario elementary schools. Thus, I outline the various statistical results I generated that extend the interpretation of school-level assessment results, and I summarize what principals, teachers, and parents at a number of "good" schools — those that perform better than schools with similar socioeconomic characteristics — told me about how their schools function.

My second goal, however, was less traditional, at least in my experience of writing as an economist: I was determined to offer my own conclusions from the research, even if it meant exposing my own biases. Thus, although I believe my conclusions follow from the results of the statistical research and from what I heard during my school visits, they cannot all be "proved" to be correct. I offer my own evaluation of the elementary school assessment process operated by EQAO, and I recommend changes where I see room for improvement. I make no claim to be an expert in how an elementary school or school board works, and I write as an outsider to much of the elementary education process. I leave it to the reader to judge the usefulness and appropriateness of this perspective.

Lessons from the Statistical Analysis

Four clear lessons can be drawn from the statistical analysis of school-level EQAO results.

First, *average test results change from year to year.* It is impossible, however, to determine whether an increase in the average across a large sample of schools is the result of an easier test, easier grading of the same test, or a general improvement in students' ability to complete the test. This does *not* mean that the assessment process is useless. If one assumes that the tests are roughly equal in difficulty and graded in a similar way from year to year, increased scores would be the result of improvements in students' ability; there is no direct evidence for or against the validity of this assumption. But even if one is unwilling to make such an assumption, the data can still be interpreted in a meaningful way.

Second, *it is possible to construct a relative ranking of schools.* A school's test results can be adjusted to take into account the socio-economic characteristics of the community from which the school draws its students, as well as the year-to-year variability in the test results. Such a relative ranking compares schools with similar composite characteristics to see if some schools have systematically better results than others over a period of years — in other words, to identify "good" schools.

Third, *the role of context — the socio-economic characteristics of the parents who send children to an elementary school — is important, but it is not the only determining factor in a school's success.* Postal codes and census data can be used to generate a multidimensional measure of context, and the list of variables that are important in a statistical sense is quite large. Yet such context variables explain perhaps 40 to 50 percent of the variation in results across schools when a four-year average of results is studied. Even allowing for the possible importance of variables omitted from this analysis, it seems certain that a school's test results depend very much on what happens at the school.

Fourth, *it is much more meaningful to average a school's test results over a number of years than to look at changes from year to year*. The results may vary substantially from year to year simply because the number of students in a class or in a grade is quite small, and better or worse scores by a handful of students can have a significant statistical effect. Averaging cancels out such effects.

In addition to these four major conclusions from the statistical analysis, one can draw two somewhat less important lessons. First, since the relative importance of context variables — particularly the relationship between test results and parents' level of education — differs between Grade 3 and Grade 6, it is important to study results from the two grades separately. Second, the rate at which students are exempted from tests varies sufficiently — from one board to another, from one school to another, and from one academic year to another — to cast serious doubt on the usefulness of any methodology, such as EQAO's Method 2, that does not take exemptions into account. Student performance should be measured using the percentage of *all* students, not just that of students who took the test.

What I Heard at Visits to "Good" Schools

As a follow-up to my statistical analysis of assessment results, I visited 13 "good" schools — schools whose results over a four-year period were usually better than 90 percent of schools with similar socio-economic community characteristics. On the basis of what I heard during my interviews with principals, teachers, and parents, "good" schools have the following characteristics.

- At "good" schools, teachers work together as a team — of particular importance in the primary division, which includes Grade 3.
- "Good" schools prepare their students thoroughly for the tests.
- In the primary division, "good" schools make use of learning resources such as mathematics manipulatives and levelled books.

- "Good" schools make good use of volunteers.
- "Good" schools with lower socio-economic characteristics rely on the principal's leadership on discipline and behaviour.
- "Good" schools have strong extra-curricular programs.
- "Good" schools communicate effectively with parents about expectations, homework, and the role of parents.
- "Good" schools generally do not sacrifice other important school activities to concentrate on improving assessment scores, although some teachers worry about reductions in activities related to the arts, music, drama, social studies, and science.

My Views of the Effectiveness of the Assessment Process

Is the EQAO assessment process "good" for Ontario's elementary school system? Such a straightforward policy question deserves an equally straightforward answer, yet one is not easy to come by. As an economist, my inclination is first to weigh the benefits and costs of the assessment process.

Benefits of the EQAO Process

The benefits of the EQAO process seem quite clear. First, although a great deal of emphasis has been placed on the reporting of overall results, the largest benefit is also the least visible: the process fleshes out the curriculum expectations for students for the end of the primary and junior divisions.[1] Without assessments, it would be more difficult for teachers to know what the expectations were for students across the province. Over time, the EQAO process has generated a reasonable clear standard of satisfactory or excellent academic work, which schools can use to identify weaknesses in students and in the coverage of their teaching and design responses to those weaknesses.

---

[1] Several persons involved in this process suggested to me that the EQAO assessments and the curriculum are not aligned, although this opinion was not widely expressed at the schools I visited. The question clearly requires further investigation. The only document I could find that addresses this issue seems to suggest that the 1998 Grade 3 assessment reasonably matched the 1998 Grade 3 curriculum (see Ireland 1998).

A second benefit, which an economist's mindset finds it easy to see, is that the EQAO process provides an external check on the organization. The buzzword is "accountability": any organization needs a mechanism to ensure that its goals are met. A business that has no customers has not met its goals. A university that attracts only weak students or whose faculty fail to obtain outside research grants has not met its goals. A hospital or nursing home whose patients die at a higher rate than expected has not met its goals — indeed, such an institution would be the subject of an investigation. One could argue that, before the EQAO assessment process began, no such external, public checks on Ontario's publicly funded elementary school system were in place. When facing a difficult or "problem" student, teachers (and university faculty members) always have some incentive to mark "too easy" — it is less trouble and less confrontational, and the problem is simply passed on to the next teacher. Now, if the assessment process reveals that the marking of a particular teacher or school varies greatly from that of EQAO, this is useful information for the teacher, the principal, the school board, and the student's parents.

A third benefit of the EQAO process is that, over time, it has provided high-quality teaching material that teachers have put to effective use, particularly in the mathematics curriculum, which is much more demanding and challenging than it used to be.

A fourth benefit is that teachers who choose to spend part of the summer as EQAO markers report this activity to be useful for their professional development, for several reasons. They saw a very large sample of student work, they saw how other teachers marked, and they saw the curriculum in action in a very broad sense. For the teachers involved, double-marking represents a significant learning opportunity, as I can personally attest from my own teaching career.

A fifth benefit is that the assessment results allow the kind of school-level evaluation of outcomes that I have undertaken in this book and the discovery of factors contributing to schools' success that are at variance with the assertions of some within the education community. In addition, my visits to "good" schools are an example of what can be done after careful analysis of the data. More visits and more research would clearly be helpful, particularly on why some schools perform very badly on assessments relative to similar schools.

Costs of the EQAO Process

The EQAO process also has costs, both monetary and nonmonetary. As I noted in Chapter 2, the monetary costs of the EQAO process are not zero, but neither are they large enough to represent a "waste" of resources on testing that would be better spent on creating thousands of new teaching or educational assistant positions, as some teachers and parents mistakenly assume. Abandoning EQAO testing would save the average 300-student elementary school just $7,500.[2] Clearly, such an amount would not begin to approach the cost of even a single staff position.

As for nonmonetary costs, one general class of costs relates to poor analysis of the school-level data. If unnecessary actions follow from a poor analysis of the data, they waste resources. Suppose, for example, that, purely by chance, a school has a poor writing result one year. If, in response, the school unnecessarily invests a great deal of time and energy in improving its students' writing, these resources are wasted. Year-over-year results contain so much noise that a rapid response to a fluctuation in an individual year is not appropriate. Teachers know this about the EQAO data — either as a mathematical property or intuitively — but it may be difficult to explain this statistical fact to parents and to the public as a whole.

A second potential cost associated with misinterpreted data involves parents' responses to results. Unless school-level results are presented in context, parents might be tempted to move in order to place their children in a school that seems "good" on the basis of absolute rankings, but that, in fact, performs poorly relative to schools with similar socio-economic characteristics. It is not clear which school would actually produce a better result for their child, but the money parents spent moving to find a "good" school would be wasted.

---

[2] For the 2001–02 academic year, the EQAO budget was approximately $50 million for 2.1 million elementary and secondary school students, or about $25 per student.

Indeed, results taken out of context might mislead not just the public but teachers as well. Some might be tempted to seek a transfer to a school with absolutely high assessment results; others, at "good" schools in communities with lower socio-economic characteristics, might become frustrated by the inability of people to understand just how uneven the playing field is and also seek to move. In either case, resources spent on teacher placement and training are wasted.

The solution to problems associated with responses to the poor analysis and presentation of EQAO assessment data is to analyze the school-level data properly — a straightforward and not very costly task — and to improve the presentation of the data and their accompanying explanation.

Another potential nonmonetary cost of EQAO testing relates to the distribution of curriculum effort. In other words, if the material stressed in the EQAO assessment (literacy and numeracy) makes it impossible to cover other equally important areas of the curriculum, as some teachers and parents fear, this must be considered a cost of the EQAO process. Does the reduced time spent on art, music, seasonal activities, science, and social studies represent a large cost? The majority of teachers I interviewed did not think so. In any event, it is my personal opinion that an elementary school should be willing to emphasize literacy and numeracy, foundations without which nothing else is possible. I might even place this "cost" as a benefit of the EQAO process.[3]

Some argue that the EQAO process imposes nonmonetary costs in terms of stress on students. At the "good" schools I visited, however, this was simply not a majority view. Doing well on the assessment meant learning to approach it in a well-prepared manner that made it a comfortable process. Again, one might even regard this not as a cost but as a significant benefit. Grade 6 students, in particular, are old enough to benefit from such a preparation process — after all, they are about to enter senior public school and high school, where this type of test situation will be a standard part of their lives for years. Perhaps a practice assessment in a low-stakes setting is a good idea.

---

[3] The best mission statement I have ever seen, and one of the few that could actually be connected to actions, is that of the Boston-Townsend School, in the Grand Erie District School Board. "We learn to read, we read to learn" seems to me to summarize concisely what should be the main goal of an elementary school.

The assessments also impose stress on teaching staff, some of which relates to how the data are analyzed and presented. At the one school where teachers refused to answer my questionnaire, substantial resentment remained, seven years after the fact, about the imposition of the EQAO process on the elementary school system. Although teachers at most schools had some legitimate complaints about the process and useful suggestions for its improvement, my perception is that most recognized its benefits.

I do wonder about the attitudes of teachers in schools I did not visit. Remember that all the schools I visited did very well on the tests and were committed to investing time and resources to prepare for them. Do teachers in other schools also find the EQAO process useful, or are those who angrily refused to respond to my queries representative of a much larger group? In any event, it is my own opinion that (to quote a popular television show) "resistance is futile." Some type of elementary school assessment is now the norm in North America. It seems very unlikely that taxpayers will be convinced to spend billions of dollars on education without some type of accountability for quality. To remove the EQAO assessment process without replacing it with another, similar process would be perceived as returning to a system without accountability.

A Concluding Comment

After weighing the costs and benefits of the EQAO process, I conclude that enough good things are coming out of the process to outweigh the bad things. I freely admit this is a subjective conclusion, not a quantitative conclusion of the type more favoured in economic analysis. I view the monetary costs as relatively small compared with the likely net benefits. And I heard enough sensible criticism of the EQAO process as it now stands to be able to offer some suggestions for improvements that are both possible and practical.

Recommendations for Improving the EQAO Process

The research I have presented in this book points to a variety of ways in which the EQAO process could be improved. I realize that my research, even though it draws on substantial statistical and analytical experience, hardly places me in the same category as someone who has spent a lifetime studying elementary schools. I know that, in suggesting improvements, I risk criticism from participants in the elementary school system, who might with some justice point to my position as an "ivory tower university professor" (see Box 8.1 for an attempt at a pre-emptive defence). Yet debate is part of the policy process. Reasonable people can disagree about the usefulness and feasibility of my suggestions, and if they encourage further discussion that leads to even one improvement, the public policy process will have been successful.

I present two sets of recommendations, one aimed at EQAO itself, the other at principals, teachers, and parents of Ontario's elementary school children.

Policy Suggestions for EQAO

My first and foremost recommendation for EQAO, following a strong and clear message from teachers at every school I visited, is to push back the date on which assessments are conducted during the academic year — as teachers told me, "the later the better." Doing so would put less pressure on teachers to complete the curriculum before the assessments are held.

Second, again following advice I received from teachers, simplify and clarify the language used in the assessments — in one teacher's memorable phrase, "avoid wonky questions."

Box 8.1: *Is There a Case for an EQAO for Universities? The View from the Ivory Tower*

Is it hypocritical for someone who sits in the proverbial university ivory tower to make recommendations about improving the elementary school system while assuming all is well in his own sector of the educational process? Universities, as with all institutions that I have ever been involved with, do some things well and others poorly, and change comes slowly and with difficulty.

Suppose Ontario decided to create an Educational Quality and Accountability Office for Universities (EQAOU) — and one might as well include colleges, too. Would such an entity be necessary? The answer is both yes and no. I would argue that it would be less necessary for universities because they are already accountable to their various granting agencies for the quality and quantity of their research. Universities also already issue some public information — albeit mostly involuntarily in response to public pressure generated by the well-known *Maclean's* magazine survey, which is a private market solution to university accountability issues. Moreover, universities compete for students on the basis of both quality of service and price, both of which are new accountability issues for the university sector. Had I been told ten years ago that my university would be competing openly for students on such bases, I would not have believed it. And if such accountability mechanisms are effective, then the case for an EQAOU is weak.

As for the case in favour of an EQAOU, I would point out that universities are not very accountable for the quality of teaching they offer undergraduates. Although universities compete for them, such students can obtain virtually no useful information on the quality of undergraduate teaching as they select which institution to attend. An EQAOU could thus be a tool for improving quality and accountability in undergraduate teaching. As is the EQAO process with teachers, an EQAOU likely would be unpopular with my professorial colleagues, but even unpopular policies can improve public outcomes. For example, the office could require universities to post teaching evaluations for lecture courses, or that the evaluations be comparable or even identical across universities, as in the case of EQAO's standardized assessment. Indeed, just such an evaluation of undergraduate teaching — the Students Evaluation of Educational Quality — is already freely available and well-respected worldwide. Clearly, there is room for public policy improvement in this area, and an EQAOU might easily accomplish that task.

Third, rename the Grade 3 and Grade 6 assessments the "Primary Assessment" and the "Junior Assessment." On the face of it, this is merely a cosmetic change, which would have the happy advantage of being cheap, but it would also take some of the pressure off Grade 3 and Grade 6 teachers by making it extremely clear that the results of the assessment are owned by the entire staff at a school.

Fourth, change the way in which the school data are presented. Here, I have three concrete suggestions:

- Abandon presenting so-called Method 2 results; they are a distraction and they increase incentives for schools to manipulate the process by which students are exempted from taking assessments.
- In addition to presenting Method 1 results and exemption rates, EQAO could present results by school in the form of three-year averages, the current three years, and the previous three years, and as deviations from the provincial average as well as in the form of levels. Such a presentation would offer a much fairer picture of improvements at a school over time and remove some of the effects of year-to-year variations in the class.
- Place the school's socio-economic context variables — that is, the school community's characteristics — in the public domain.

Fifth, do a better job of explaining to the public, and even to teachers, the value to the educational process of responses at the school level to long-term, multi-year weaknesses in specific areas of the curriculum as identified by the assessments. EQAO should also stress the value for accountability in having an outside marker for a *small* portion of a student's work.

Finally, work harder to clarify the actual monetary costs of EQAO's operation, since teachers and parents appear to believe that those costs are much higher than they really are.

Policy Suggestions for Principals, Teachers, and Parents

My suggestions for principals, teachers, and parents could be viewed in two ways. First, they could be considered as the salient lessons

I learned at "good" schools turned into policy ideas. Second, they could be seen as questions that parents could ask about their particular school: is this activity or concept in place at the school, and if not, why not?

I present the following suggestions approximately in their order of importance.

First, schools should recognize the enormous value of having teachers work as a team in both the primary and junior divisions. Although this may be more difficult at a larger school where there are double or even triple classes across grades, the needs of both teachers and students are better met within such a framework.

Second, schools should ensure that teachers have the learning resources they need. As noted previously, teachers at "good" schools cited levelled books and mathematics manipulatives as particularly helpful tools.

Third, schools should recognize the important role volunteers can play. Moreover, it is not enough simply to have volunteers; volunteers must be trained and directed to useful tasks — primary programs in reading, for example — in order to make a real contribution to assessment results. Volunteers need to feel their time is used well and focused on the needs of students. Parent volunteers probably should not be placed in the same classrooms as their children, otherwise they could be a distraction; moreover, such volunteers should be encouraged to think of themselves as helping the school community as a whole, not just their own child.

Fourth, schools in communities with lower socio-economic characteristics, where parent volunteer resources may be more limited, should make greater use of volunteers from local high schools, colleges, and universities and from among retired people in the community.

Fifth, parents on school councils should try to find a balance between being "supportive" and being "demanding." Ask the school to explain specific weaknesses in assessment results that show up over a number of years, but understand that a large fluctuation from one year to the next is a common statistical feature of the data.

Sixth, both parents and teachers should ensure that students approach assessments as a low-key, calm activity and with confidence that they have been well prepared.

Seventh, schools should communicate effectively with parents on a regular basis.

Eighth, parents should not be obsessed with test results. "Good" schools are those with balance and strong extracurricular activities that require parental input and support. Teachers appreciate that kind of support.

Finally, schools should give added importance to daily physical education, despite the difficulty of scheduling such activity in the school day. The benefits, at both the primary and junior levels, would be worthwhile.

Questions for School Boards

In thinking of improvements arising from my research, I found that decisionmakers at local school boards presented the most difficulty. School boards face special problems. For example, management is tightly constrained by collective agreements, meaning that school boards are unlikely to be able to change the rules and allow principals more autonomy in their selection of staff as they attempt to build effective teams. In addition, major financial decisions are often taken (by government) that are beyond the direct influence of school board decisionmakers.

For this group, then, I would be more confident phrasing my ideas as questions than as suggestions, roughly in order of their apparent practicality.

First, is the "normal" length of time a principal spends at an elementary school appropriate? From my school visits, I sensed strongly that, given the much slower turnover of teaching staff and the time it takes to build effective teams of teachers, principals are not given the opportunity to stay long enough to have much impact. In addition, I learned from my meetings with a number of school councils that, although school boards believe, in theory, that councils should play an active consultative role in the process of selecting principals, most pay little more than lip-service to the idea. If a proposed new principal had to offer an explicit statement of his or her vision for the school and was given enough time to implement that vision, would more schools develop effective teams of primary and junior teachers?

Second, are boards spending enough money on professional development? Teachers at the schools I visited appreciated high-quality professional development, and felt it was money well spent.

Third, should school boards ensure that every school contains at least one teacher — and preferably two, one for Grade 3 and one for Grade 6 — with experience marking EQAO assessments? It was clear from my school visits that such teachers served as a useful resource for other teachers. How could a board implement such a suggestion? Perhaps it could offer a monetary incentive to encourage teachers with EQAO marking experience to transfer to schools with no such people on staff. A monetary incentive might also encourage younger, lower-paid teachers to undertake EQAO marking during the summer. Perhaps EQAO marking could be considered useful experience when boards offer new teachers permanent contracts.

Fourth, should there be a limit on the amount of time a teacher stays at the same school? Although there are benefits to a school's having experienced teachers who have worked together for a long time, those fortunate enough to find positions in schools serving communities with high socio-economic characteristics are understandably reluctant to move to more difficult teaching positions in schools in less desirable neighbourhoods. At the same time, the less desirable schools have the least parental support and are usually staffed with the least experienced teachers. To an outsider, this allocation of teaching experience and talent does not seem best for the children in the system. Could some kind of mandatory movement of teachers help spread experienced teachers around to a larger variety of schools? Would this help principals build effective teams of teachers?

Fifth, how do boards allocate resources among schools in communities with very different socio-economic characteristics? Do more resources find their way into weaker schools and if so, by what process? Some boards I visited concentrated their early literacy efforts in schools with higher needs, but have these extra resources made a difference? It is my hope that the analysis in this book will help school board decisionmakers determine the answer to that question. I recognize, of course, that this type of policy recommendation walks a knife edge. Schools in communities with higher socio-economic characteristics need to be well equipped and well funded to remain attractive to affluent parents; if sufficiently large resources are diverted from such schools to poorer schools, affluent parents might be more likely to consider private alternatives for their children, with consequent loss of significant and important political support by the publicly financed system. The balance point would be hard to find. One could, for example, fund all schools equally, in the sense of providing the same number of dollars per student or maintaining similar ratios of students to teachers. It is reasonably clear, however, that such a "fair" approach would not generate "equality of opportunity" for students whose relative lack of home resources puts them at a disadvantage compared with children from more affluent homes. One of this study's findings is that context variables do make a difference to students' results on standardized assessments — all students are *not* equally good at reading, writing, and mathematics, regardless of their socio-economic backgrounds. It seems to me that students from less advantaged backgrounds might benefit greatly from the transfer of some resources from "rich" schools to "poor" schools within a board — in the form of, for example, smaller classes or more Educational Assistant resources in weaker schools. A large "rich" school with minimal discipline problems might not need a vice-principal, but a smaller, "poor" school might require one desperately. In short, although such strategies may already be in place in some schools, boards need to think more about how to allocate their resources more effectively, and take appropriate action.

A Final Note

There is no doubt that the introduction of the EQAO assessment process has had an enormous effect on Ontario's elementary schools. In this book, I have attempted to interpret in a variety of ways the vast amount of data the assessment process has provided from the 1998–99 through 2001–02 academic years. Using those assessment results and substantial census information on the socio-economic characteristics of school communities, I have described a method of identifying schools that do better on the tests than other schools in communities with similar characteristics. I visited a number these "good" schools, and I have summarized what principals, teachers, and parents told me about how their schools operate. Finally, I have suggested some possible areas of improvement and ventured some questions that participants in the EQAO process might wish to consider.

I hope administrators, teachers, parents, and, ultimately, students find this contribution to be useful.

About the Author

David Johnson is Professor of Economics at the School of Business and Economics at Wilfrid Laurier University. He was educated in the Ontario public school system in Hamilton and London, then at the University of Toronto, the University of Western Ontario, and Harvard University.

To be upfront on one issue, both Johnson children attended public schools in the Waterloo Region District School Board and, for a year, taxpayer-supported schools in the Cambridgeshire (England) Local Education Authority to the end of Grade 6. After Grade 6, both children attended Rockway Mennonite Collegiate in Kitchener, a school operated by the Mennonite Conference of Eastern Canada without direct taxpayer support. The decision by the Johnson children and their parents to support Rockway Mennonite Collegiate was not a reflection of the quality of education offered at the Waterloo Region District School Board. Rather, their decision reflected the approach Rockway is able to take to meet the whole needs of a student — spiritual, physical, and intellectual — in the context of the Mennonite church community.

Most of Dr. Johnson's previous research is in the areas of macroeconomics, monetary economics, and international finance. However, he became interested in the processing and interpretation of elementary school assessment data. After identifying schools that outperform similar schools, it became impossible for him to resist the desire to visit these schools to find out more about them.

Research in economics, Dr. Johnson says, is about asking questions that can contribute to the public policy debate and to better outcomes.

Members of the C.D. Howe Institute[*]

Corporate, Association, Foundation, and University Members

Accenture Inc.
Acklands-Grainger Inc.
Advocis
AGF Management Limited
Alcan Inc.
Allstream Inc.
The American Chamber of Commerce in Canada
AON Consulting
ARC Financial Corporation
ARC Resources Ltd.
Association of Canadian Pension Management
Association of International Automobile Manufacturers of Canada
ATB Financial
Bank of Nova Scotia
Barrick Gold Corporation
BCE Inc.
Belkorp Industries Inc.
Bell Canada
Bennett Jones
The Birks Family Foundation
BMO Financial Group
Boardwalk Equities Inc.
Bombardier Inc.
Borden Ladner Gervais LLP
BP Canada Energy Company
Brascan Corporation
British Columbia Hydro and Power Authority
British Columbia Securities Commission
Brookfield Properties Corp.
Bruce Power L.P.
Burgundy Asset Management
Burlington Resources Canada Ltd.
Burnco Rock Products
Business Council of British Columbia
Business Development Bank of Canada
Cadillac Fairview Corp. Ltd.
CAE Inc.
Cameco Corporation
Campbell Strategies
Canada Deposit Insurance Corporation
Canada Overseas Investments Limited
Canada Post Corporation
Canadian Association of Petroleum Producers
Canadian Association of Recycling Industries
Canadian Bankers Association
Canadian Chamber of Commerce
Canadian Chemical Producers' Association
Canadian Energy Pipeline Association
Canadian Federation of Independent Business

---

[*] The views expressed in this publication are those of the author, and do not necessarily represent the opinions of the Institute's Board of Directors or members.

Canadian Finance & Leasing Association
Canadian Imperial Bank of Commerce
The Canadian Institute of Chartered Accountants
Canadian Life and Health Insurance Association Inc.
Canadian Manufacturers & Exporters
Canadian Oil sands Limited
Canadian Pacific Railway
Canadian Tax Foundation
Canadian Tire Corporation, Limited
Canadian Western Bank
Canam Group Inc.
Candor Investments Ltd.
CanWest Global Communications Corp.
Cargill Limited
Carleton University
CHC Helicopter Corporation
Clairvest Group Inc.
CMA Holdings Inc.
CN
Cogeco Cable Inc.
CompCorp
Concordia University
Credit Union Central of British Columbia
Deloitte & Touche LLP
Desjardins Ducharme Stein Monast, General Partnership
Deutsche Bank AG, Canada Branch
Dofasco Inc.
Domtar Inc.
Donner Canadian Foundation
Duke Energy Gas Transmission Canada
Economap Inc.
Economic Development Edmonton
Edco Financial Holdings Ltd.
E.I. du Pont Canada Inc.
E-L Financial Corporation Limited
Elk Valley Coal Corp.
Emera Inc.
Enbridge
EnCana Corporation
Energy Council of Canada
Ensign Resource Service Group Inc.
Ensis Growth Fund Inc.
EPCOR
Ernst & Young LLP
Export Development Canada
Fairmont Hotels and Resorts
Falconbridge Limited
Fednav Limited
Fidelity Investments
Finning International Inc.
Ford Motor Company of Canada, Limited
Forest Products Association of Canada
Four Halls Inc.
Four Seasons Hotels Limited
Fraser Milner Casgrain LLP
Fraser Papers Inc.
Gaz Métro
General Electric Canada Inc.
George Weston Limited
Gibson Energy Ltd.
Gluskin Sheff + Associates Inc.
Goal Group of Companies
Goldberg Group
Goodmans LLP
Greater Toronto Airports Authority
The Great-West Life Assurance Company, London Life Insurance Company and Canada Life
The Harold Crabtree Foundation
Harvard Developments Inc., A Hill Company
HET Corporation
Hill & Knowlton
Historica Foundation
Hollinger Inc.
Honeywell Canada
HSBC Bank Canada
HSD Partners Inc.

Hydro-Québec Production
IBM Canada Ltd.
Imperial Oil Limited
Imperial Tobacco Canada Limited
Inco Limited
Industrial Alliance Life Insurance Company
The Institute for Competitiveness and Prosperity
Insurance Bureau of Canada
Investment Dealers Association of Canada
The Investment Funds Institute of Canada
Investors Group
IPSCO Inc.
J.D. Irving Limited
J.P. Morgan Securities Canada Inc.
Jackman Foundation
Jarislowsky, Fraser Limited
John Dobson Foundation
KPMG LLP
La Jolla Resources International Ltd.
Laval University
Lehigh Inland Cement Limited
Lockerbie & Hole Inc.
Lombard Canada Ltd.
Maclab Enterprises
Magellan Aerospace Limited
Manulife Financial
Marsh Canada Limited
Masonite International Corporation
Max Bell Foundation
McCarthy Tétrault LLP
McGill University
McMaster University
McMillan Binch LLP
Mega Bloks Inc.
Mercer Human Resource Consulting Limited
Merrill Lynch Canada Inc.
Monarch Construction Limited
Monitor Group
Moosehead Breweries Limited
Morneau Sobeco
Mount Saint Vincent University
N M Rothschild & Sons Canada Limited
National Bank of Canada
National Hockey League Players' Association
National Leasing
New Brunswick Power Corporation
Nexen Inc.
Noranda Inc.
Norbord Inc.
NorskeCanada
Northam Realty Advisors Limited
NOVA Chemicals Corporation
Onex Corporation
Ontario Energy Board
Ontario Hospital Association
Ontario Medical Association
Ontario Municipal Employees Retirement System (OMERS)
Ontario Power Generation Inc.
Ontario Teachers' Pension Plan
Osler, Hoskin & Harcourt LLP
Penn West Petroleum Ltd.
Petro-Canada
Pfizer Canada Inc.
Phillips, Hager & North Investment Management Ltd.
Pirie Foundation
Power Corporation of Canada
Pratt & Whitney Canada
PricewaterhouseCoopers LLP
Procter & Gamble Inc.
Property and Casualty Insurance Compensation Corporation
Provident Energy Ltd.
Queen's University
The Railway Association of Canada
RBC Capital Markets
RBC Financial Group

Régie des rentes du Québec
Rogers Communications Inc.
Ryerson University
Samuel Son and Co. Limited
Saskatchewan Wheat Pool
Saskatoon & District Chamber of Commerce
Sceptre Investment Counsel
SGI
Shell Canada Limited
Sherritt International Corporation
Simon Fraser University
SNC Lavalin Group Inc.
Sun Life Financial Inc.
Suncor Energy Inc.
Syncrude Canada Limited
Tax Executive Institute
TD Bank Financial Group
Teck Cominco Limited
Tembec Inc.
The Thomson Corporation
The Toronto Board of Trade
Torstar Corporation
Torys LLP
TransAlta Corporation
TransCanada Corporation
Transcontinental Inc.
UBS Global Asset Management (Canada) Co.
Université de Montréal
Université de Sherbrooke
University College of the Fraser Valley
University of Alberta
University of British Columbia
The University of Calgary
The University of Lethbridge
University of Manitoba
University of Ottawa
University of Regina
University of Saskatchewan
University of Toronto
University of Waterloo
The University of Western Ontario
Via Rail Canada Inc.
Watson Wyatt Worldwide
Weston Forest Corporation

Individual Members

Keith P. Ambachtsheer
E. James Arnett, Q. C., of Counsel
Harry G. Baumann
David E. Bond
R.A.N. Bonnycastle
Jack Brodsky
Pierre Brunet
Robert Buttke
Robert C. Caldwell
Ken Chapman
Ben Cherniavsky
Kenneth Christoffel
Jack Cockwell
Marshall A. Cohen
E. Kendall Cork
Marcel Côté
Glen Cronkwright
Paul R. Curley
Thomas P. d'Aquino
Catherine A. Delaney
Steven Devries
Wendy Dobson
John F. Eckert
William F. Empey
Morrey M. Ewing
John T. Ferguson
Aaron Fish
Jim D. Fleck
Julia Foster
Robert Foster

Christine Girvan
Peter Goring
John A.G. Grant
James K. Gray
Roger Guillemette
John Haag
Geoffrey Hale
Victor Harding
H. Clifford Hatch
G.R. Heffernan
Lawrence Herman
William L. Holt
H. Douglas Hunter
Lou Hyndman, Q.C.
Joseph Kruger II
W. Edwin Jarmain
Robert Johnstone
B. Lynn Jones
John A. Kazanjian
James T. Kiernan
Frank F. Kolb
R. John Lawrence
Jacques A. Lefebvre
David A. Leslie, FCA
David Lindsay
Anna Loncar
J.W. (Wes) MacAleer
William A. Macdonald, Q.C.
James P. McIlroy
Robert M. MacIntosh
Frank McKenna
John D. McNeil
Georg Marais
Bruce E. Mintz
William Molson
Russell J. Morrison
F.W. Orde Morton
John P. Mulvihill
Edward P. Neufeld
William Orovan
James S. Palmer, C.M., ADE, Q.C.
Andy Perry
Donald S. Reimer
Barrie D. Rose
J. Nicholas Ross, C.A.
Guylaine Saucier
Hugh D. Segal
Lindsay K. Shaddy
Gordon Sharwood
Philip Spencer, Q.C.
Wayne Steadman
Morley Stock
Harry Swain
Christopher Sweeney
Thomas H.B. Symons
Frederick H. Telmer
Jack H. Warren
David J.S. Winfield
Alfred G. Wirth
Adam H. Zimmerman

Honorary Members

John Crispo
Grant L. Reuber

Recent Publications of the C.D. Howe Institute

February 2005	Poschmann, Finn, and William B.P. Robson. "Restoring Purpose: A Shadow Federal Budget for 2005." *C.D. Howe Institute Commentary* 208.
January 2005	Chen, Duanjie, and Jack M. Mintz. "How to Become Seductive: Make Canada More Investment-Friendly." *C.D. Howe Institute e-brief.*
January 2005	Guillemette, Yvan. "School Enrolment Is Down; Spending Is Up: What's Wrong with This Picture?" *C.D. Howe Institute e-brief.*
December 2004	Wright, David S. "Managing Global Crises — and the U.S. Colossus." *C.D. Howe Institute Commentary* 207.
December 2004	Robson, William B.P., and Danielle Goldfarb. "Tools for Workers: How Canada Is Faring in the Competition for Capital Investment." *C.D. Howe Institute Backgrounder* 87.
December 2004	Mintz, Jack M., and Tom Roberts. "Holes in the Road to Consensus: The Infrastructure Deficit — How Much and Why?" *C.D. Howe Institute e-brief.*
December 2004	Guillemette, Yvan, David Laidler, and William B.P. Robson. "The Real Reasons for the Canadian Dollar's Power Trip — and What Not to Do about It." *C.D. Howe Institute e-brief.*
November 2004	Goldfarb, Danielle. "How Canada Wins from Global Services Outsourcing." *C.D. Howe Institute Commentary* 206.
November 2004	Poschmann, Finn, and William B.P. Robson. "Saving's Grace: A Framework to Promote Financial Independence for Low-Income Canadians. *C.D. Howe Institute Backgrounder* 86.
November 2004	Gotlieb, Allan. *Romanticism and Realism in Canada's Foreign Policy.* Benefactors Lecture, 2004.
October 2004	Mintz, Jack M., and Finn Poschmann. "Follow the Cash: Changing Equalization to Promote Sound Budgeting and Prosperity." *C.D. Howe Institute Backgrounder* 85.

Signposts of Success

Interpreting Ontario's Elementary School Test Scores

David Johnson

In the 1996-97 academic year, Ontario introduced standardized testing in reading, writing and mathematics for all elementary school students in the province, at first for those in Grade 3 and later extended to those in Grade 6 as well. Considerable controversy surrounds these tests, particularly when the results are used t create report cards on each school's effectiveness. Many critics argue that school ranking are based, not on the school's relative success in teaching students, but on the socio-economic characteristics of the community from which the school draws its students, thus unfairly giving lower rankings to schools in poorer neighbourhoods.

Are school report cards, in fact, useful? Are schools that contain students with higher average test results actually better schools? If better student test results are not an appropriate measure of a school's success, what is? Is the provincewide elementary assessment program itself useful?

In this trailblazing book, David Johnson shows that socio-economic factors do not explain all the differences in school rankings on test scores - that principals' managerial talents, the quality of teaching, and the resources available to the school also affect students' achievement scores. This is good news for educators, because it means they can make a difference in the classroom.

David Johnson is Professor of Economics at the School of Business and Economics, Wilfrid Laurier University. Most of Dr. Johnson's research has been in the areas of macroeconomics, monetary economics, and international finance, but he became interested in the processing and interpretation of elementary school assessment data. After identifying schools that outperform similar schools, it became impossible for him to resist the desire to visit them to find out what made them better. Research in economics, Dr. Johnson says, is about asking questions that can contribute to the public policy debate and to better outcomes.

ISSN 0832-7912
ISBN 0-88806-649-X
$21.95

ISBN 0-88806-649-X
9 780888 066497